CHAPTER 7

Production Function

1. What is a production function? What are irs features?

Production is the transformation of inputs into output which adds to their utility and value. How much output is produced depends on the level or quantity of inputs used per unit of time. In short, output depends on inputs or output is a function of inputs. This functional relationship between inputs and output is called production function in economics. More precisely, **a production function is a physical/technical functional relationship between output of a good and the inputs required to make that good which specifies the maximum output that can be produced with given inputs for a given level of technology.**

Four things need to be noted regarding a production function:

1. A production function refers to the relation between inputs, expressed in physical terms, and output, also expressed in physical units: prices of inputs or output do not enter in to production function.
2. A production function is a functional relationship or the relationship of dependence between inputs and output where output depends on inputs. Thus, output is the dependent variable and inputs are independent variables.
3. A production function gives maximum output possible with a given set of inputs, assuming that the inputs are used efficiently.
4. A production function is defined for a given technology. It is the technological knowledge or scientific know-how that determines the maximum levels of output that can be produced using different combinations of inputs. If the technology improves, the maximum levels or quantities of output obtained for the given input combinations increase. We then have a new production function.

In formal notation the production function has the general form:

$$Q = f(L, N, K, R)T$$

Where Q is output and L, K, N, R and T are inputs with L denoting labour, N land, K capital, R raw materials and T technology. Existence of T outside the parentheses means input-output relation is defined for a given state of technology. The letter f tells us that Q is a function of inputs, that is, the inputs determine the Q.

Because every production function is defined for a given state of technology (shown by keeping the T outside the parentheses) we can omit the symbol T to write the production function in the following form:

$$Q = f(L, N, K, R)$$

It has been observed that raw materials (R) have a constant relation to output at all levels of production. For example, the material required for a certain type of computer is constant, irrespective of the number of computers produced. This allows the exclusion of raw materials from the production function to give us

$$Q = f(L, N, K,)$$

For the sake of simplicity and for making it possible to use a two dimensional diagram for representing a production function, input land (N) is lumped together with machinery and equipment in capital (K). This leaves us with the following final expression of a production function:

$$Q = f(L, K)$$

This reads as 'the output of a good is a function of the amount of labour and capital used, and says that by using L amount of labour and K amount of capital, we can at most produce Q amount of output.'

2. What are different types of production functions?

Ans. There are two broad classifications of production functions. One is on the basis of variability or otherwise of factor ratios and the other is on the basis of possibility of variations in inputs. On former basis, productions functions are classified as constant proportions type production and variable proportions type production. On the basis of later classification production functions are classified into short run and long run productions.

1. Constant Proportions Type Production Function

Constant proportions type production function is the production function in which all inputs continue to be used in the same, constant or fixed proportion, no matter what the level of output. In this type of production function quantities used of all inputs change but in a constant ratio. This is also called fixed proportions type production function.

When the amount of all inputs is changed in a constant ratio, we are actually scaling[1] the amount of all inputs up by some constant factor, for example, twice or thrice as much of all inputs. When this happens, scale of production is said to have changed. Thus, scale of production refers to the production capacity of a firm: it is related to some constant change in the amount of all the inputs used in a production process.

2. Variable Proportions Type Production Function

Variable proportions type production function is the production function in which input ratio tends to change corresponding to different levels of output. In this type of production function, quantity used of some inputs changes while that of others remains fixed.

Short Run and Long Run

Now the question is '*why are some inputs variable and others fixed*?' This question can be answered by considering the time required to change the quantity of inputs.

1. Scaling means expanding or shrinking something by some factor.

In some situations, when there is need to increase output, a producer cannot increase the quantity used of all inputs immediately. For instance, if the demand for a certain brand of computers rises suddenly, the producer cannot immediately obtain new land, construct new buildings and purchase new (sophisticated) machinery: time is required to make arrangements for such inputs. But the producer can purchase some more amounts of raw materials and hire some more workers at a short notice. He can also ask existing workers to work overtime. Thus, the firm is able to change the quantities used of some inputs but not others. This is the reason for some inputs being variable and others being fixed.

Keeping in mind the fact that all the inputs cannot be varied with the same ease, we can think of two time periods in the theory of production: short run and long-run.

Short run is the period of time which is not long enough to change the quantities of all inputs used by the firm when it needs to increase the level of output, and the only option available to it is to utilise existing plant[1] more intensively by increasing the quantities of inputs which can be changed at a short notice. Those inputs the quantity used of which can be changed or varied over the time-period under consideration are called variable inputs and those inputs the quantity used of which cannot be changed, that is, it remains fixed, are called fixed inputs. The inputs which are usually variable in the short run include labour and raw materials. The inputs which remain fixed are generally those which are indivisible. Indivisibility of factors means that some factors are available in some minimum sizes; they are not available below that minimum level and cannot be divided into smaller sizes to suit a smaller level of production. Examples of such factors include machinery, buildings (also called factory space), land, managerial staff etc.

Long run is the period of time which is long enough to change the quantities used of all inputs and the firm has the option to change the plant size. Since the amount of all inputs used in a production process, which determines the production capacity of the firm, is called the scale of production, we can also define the long run as the period of time in which scale of production can be changed. Thus, in the long run all inputs are variable and no input is fixed. Hence there is no distinction of fixed and variable inputs in the long run; a machine employed for production may be considered as a fixed input in the short run, but given a sufficiently long period, the number of machines can be varied in accordance with the requirements of the scale of production, thus making it a variable input.

For any particular production process, long run generally refers to a longer period of time. For different production processes or industries, the time span to be categorised as short run or long run may be different. For example, it takes three or more years to build a new hydro electric power station but it will take only a few weeks to arrange for a new driller. For a drilling company one month is a long period because it takes only a few weeks for it to arrange for a new driller. But for a power company, three years are a short run because it cannot build a new complete power house in a shorter time period.

The above discussion leads us to the conclusion that time in economics is a functional concept and not the usual calendar concept which will regard one month a short time period and three years a long time period. This means, in production theory, short run and long run are not defined in terms of days, weeks, months or years. Instead, we categorise a time period as a short run or long run simply by looking at whether all the inputs can be varied or not; if all inputs can be varied, it is a long run, but if all inputs cannot be varied (at least one input remains fixed) then it is a short run.

1. 'Plant' in this context means a factory, mill, or other assemblage of connected productive facilities located on a single site.

From the above discussion it follows that

- Constant proportions type production function is a long run concept: it is an input-output relation in which all factor inputs are variable and no input is fixed. This is possible in the long run only.
- Variable proportions type production function is a short run concept: it is an input-output relation in which at least one factor remains fixed and others are variable. A firm can change its level of output by changing its variable inputs only and not fixed inputs. This happens in short run.

Whether we are considering short run or long run, one thing has to be borne in mind. That is, the technology or technique of production at the disposal of producer remains fixed or unchanged. The period of time in which technology changes or new techniques of production are introduced is called very long run. For example, introduction of computers in place of typewriters. In the long run firms produce with the techniques currently available. In the very-long run production techniques change. This means the production function itself changes so that same level of inputs produces more output.

3. What are laws of production?

Ans. In general, an economic law is a statement of the general tendency of an economic variable: it gives us information regarding the general behaviour of an economic variable. In other words, an economic law is a statement generalising the tendency of a cause to produce some result.

Laws of production, thus, are the statements of the general tendencies of change in outputs of goods as a result of change in inputs. In other words, laws of production are the statements specifying the nature of change in output of a good as inputs change and describe technically the possible ways of increasing the level of production.

There are two types of laws of production: one, law of variable proportions, and two, law of constant proportions

(*i*) Law of Variable Proportions

Law of variable proportions gives the general tendency of change in output of a good as inputs change in short run. Since, in short run a producer cannot change the quantities used of all inputs- quantities used of only some inputs can be changed while those of others cannot be- the ratio in which factor inputs are combined is different corresponding to different levels of output. It is for this reason that this law is called the law of variable (changing) proportions. So we can say that the law of variable proportions gives the general tendency of output of a good as factor proportions change in the short run.

For the sake of simplification we assume that only one factor is variable and all others are fixed so that we are able to predict the effect of a change in this input on the output of the good. For this reason the law of variable proportions is often referred to as the returns

to a factor. In this sense this law gives the general behaviour of output of a good as the quantity used of a single input is changed. Thus, returns to a factor means the general tendency of output of a good in response to the change in the quantity of one input used with the fixed quantities of other inputs.

(ii) Law of Constant Proportions

Law of constant proportions gives the general tendency of change in output of a good, in long-run, as quantities of all inputs change simultaneously. There can be two situations when a producer increases all inputs simultaneously.

One, when the quantities of inputs are increased in the same proportion so that the ratio of inputs remains constant at all levels of output and the other, when quantities of inputs are increased in different proportions so that the ratio of inputs is different for different levels of output. Traditional theory of production concentrates on the first case, that is, the case of constant proportions. That is why this law is often referred to as the law of constant or fixed proportions. Thus, the law of constant proportions gives the general tendency of change in output of a good, in long-run, as quantities of all inputs change simultaneously in a fixed proportion.

This law is also called the returns to scale. The reason is that the change in all inputs at a constant ratio is called the change in scale; the changes in output as all factors change by the same proportion are referred to as returns to scale; this law gives the effects of scale relationships.

In short, law of constant proportions gives the rate of increase in output, in long-run, in response to a proportionate increase in all inputs. . In other words, this law gives the general behaviour of output of a good as the quantities used of all inputs are changed at a constant ratio.

4. Explain the concepts of total, average and marginal products.

Ans. Total, average and marginal products are the three important production concepts of short run which are very helpful in understanding the law of variable proportions. Actually, these three production concepts are three different ways of looking at how output varies as the quantity of variable input is varied.

Let's first understand the meaning of these concepts. (An important point to be noted is that all the three production concepts are physical concepts, that is, they measure production of commodities in physical units like kilograms of wheat, number of mobile sets, meters of cloth etc, and not in monetary units).

Total Product:

Total product, also called total physical product, gives the total amount of output produced during some period of time by all the inputs used by the firm, that is, both fixed and variable. If quantity used of any one input, say labour, is varied, keeping the quantities of other inputs unchanged, the total product will change as the variable input changes.

How will total product respond to a change in the variable input? This question is answered by the law of variable proportions. The law says that as the quantity of a variable input is varied, holding other input quantities constant, TP will initially increase at an increasing rate (that is, successive increments will be greater than preceding ones). After some time the rate of increase in TP will change and it will increase at a diminishing or decreasing rate (that is, the successive increments will be smaller than preceding ones). But this increasing trend in TP will not last forever: sooner or later a stage will come when TP will stop increasing with increases in variable input. If variable input is increased further, TP, instead of increasing, will begin to decrease.

But why does TP respond to changes in variable input in the manner described above? The answer is given by marginal product of variable input –another important production concept which we are going to discuss shortly.

Average Product

Average product is the output per unit of variable input. AP is obtained by dividing the TP at some point by the number of variable inputs corresponding to that level of total product. As the quantity of variable input is increased, average product first increases and then decreases. This behaviour of AP is governed by the behaviour of TP because AP is obtained by dividing the TP by the corresponding number of variable inputs. Remember that average product of labour is a common measure of labour productivity.

Marginal Product

Marginal product is the addition to the total product by the use of an additional unit of variable input or simply, the output obtained from using an additional unit of a variable input. As the quantity of variable input is increased marginal product first increases, reaches a maximum, then decreases and finally it becomes negative. (We will give a detailed explanation for this behaviour of MP at a later stage).

The calculation of these three production concepts is shown in the table given below.

Column I of the table gives number of units of fixed input, (which we have assumed to be capital), used by the producer. Since we are dealing with the short run and because we have assumed capital to be the fixed input, its quantity used will not change: it will remain fixed at 5 units whatever the level of output.

Column II of the table gives number of units of variable input, (in our case labour), used by the producer in the production process. If the producer desires to increase the level of output, he can do so by varying the quantity of labour (variable input) only so that its quantity increases as more output is produced.

Units of capital (i)	*Units of labour* (ii)	*Total product* (iii)	*Average product* (iv)	*Marginal product* (v)
5	0	0	-	-
5	1	20	20	20
5	2	80	40	60
5	3	175	58	95
5	4	300	75	125
5	5	440	88	140

Units of capital (i)	*Units of labour* (ii)	*Total product* (iii)	*Average product* (iv)	*Marginal product* (v)
5	6	575	95.8	135
5	7	685	97.8	110
5	8	770	96.2	85
5	9	830	92	60
5	10	875	87.5	45
5	11	910	82.7	35
5	12	930	77.5	20
5	13	940	72.3	10
5	14	940	67	0
5	15	930	62	– 10
5	16	910	56.8	– 20

Column III of the table gives total product. As mentioned above, total product is the quantity of output produced by all the inputs. It means TP is the result of the combined effort of all the inputs used by the firm. The first line of the table shows that total product is zero. This point is to be noted; although the firm is using 5 units of capital, nothing comes out. The explanation for this is that production is the result of the combined effort of all the inputs used. At this stage the firm is using only one input- capital, and no labour. Since capital cannot do anything on its own; someone is needed to start and operate the machines and equipment, nothing is produced. Production requires the combined efforts of both labour and capital.

When the firm starts employing labour input, output begins to come out. With the employment of first unit of labour along with 5 units of capital, 20 units of output are produced. This is the total product derived from 5 units of capital and 1 unit of labour. As the producer increases the units of labour, holding the units of capital fixed at 5 units, total product goes on increasing, as is clear from the column III of the table. But this increase is not unlimited. Up to the employment of 13 units of labour TP is increasing continuously. But the increase in TP stops here. With the employment of 14th unit of labour, TP remains fixed at 940 units which is the same as that produced with 13 units of labour. When the employment of labour is further increased, output begins to decline.

Also note that TP first increases at an increasing rate. This happens till the employment of 5 units of labour. Beyond this point and up to the use of 13 units of labour, TP increases but at a diminishing rate. It then stops to increase and finally begins to fall. Thus, TP is maximum at the employment of 13th and 14th units of labour.

Column IV of the table gives average product. Average product is the total product per unit of variable input or simply total product divided by the number of units of variable input used. That is,

$$AP = TP/L$$

Here L is the units of labour or variable input.

First line of the table shows nothing for AP because TP is zero. When one unit of labour is used, TP is 20 units. So, AP = 20/1 = 20. When two units of labour are used, AP of labour is 80/2 = 40 and so on. The point to be noted about AP is that it first increases, reaches a maximum and ultimately it too starts declining. In the table, AP rises up to the employment of 7 units of labour (here AP = 685/7 = 97.8). Beyond 7th unit of labour, it falls. Although AP will fall continuously beyond this point, it will never become zero. Zero AP means nothing is produced at all.

Column V of the table gives marginal product. Marginal product is the output obtained from using an additional unit of a variable input. There are three ways to calculate marginal product. These are

1. MP of nth unit of labour = TP of n units of labour – TP of $n-1$ units of labour, or

MP of nth unit of labour = $TP_n - TP_{n-1}$

Example : Let n = 4, so

MP of 4th unit = TP of 4 units of labour – TP of (4 – 1 = 3) units of labour

MP of 4th unit of labour = 300 – 175

= 125 units of output

2. MP of n + 1th unit of labour = TP of n + 1 units of labour – TP of n units of labour,

or MP of n + 1th unit of labour = $TP_{n+1} - TP_n$

Example : Let n = 3, so

MP of (3 + 1 = 4)th unit of labour

= TP of (3 + 1 = 4) units of labour – TP of 3 units of labour

MP of 4th unit of labour = 300 – 175

= 125 units of output

3. MP = $\Delta TP / \Delta L$

Here Δ (delta) means change in, and L means number of units of variable input (here labour) employed.

This formula is based on the following definition of marginal product:

Marginal product is the change in total product resulting from a small (one unit) change in the quantity used of variable input.

Example : When the firm increases the quantity used of the variable input (labour) from 3 to 4 units, ΔL is 1 and because of this, total product increases from 175 to 300 units, so ΔTP is 125.

Hence $$MP = \frac{\Delta TP}{\Delta L} = \frac{125}{1} = 125 \text{ units}$$

It should be noted that marginal product first increases, reaches a maximum and then starts declining. Marginal product can become zero and may even become negative. Positive marginal product, that is, MP > 0, means that the employment of additional units of the variable input is making some addition to the total product. Zero MP means that the related unit of the variable input is making no contribution to the production of the commodity. Finally, negative MP, that is, MP < 0, means instead of making some positive contribution to the production of the good, the additional units of the variable input cause reduction in the contribution made by other units.

5. Explain the behavior of different production concepts with the help of production curves.

Ans. Diagrammatic representation of the three production concepts will give us three product curves, viz, total product curve, average product curve and marginal product curve.

Total Product Curve

Total product curve is obtained by plotting the total product against the number of units of variable input. This plot gives us a curve of the following shape.

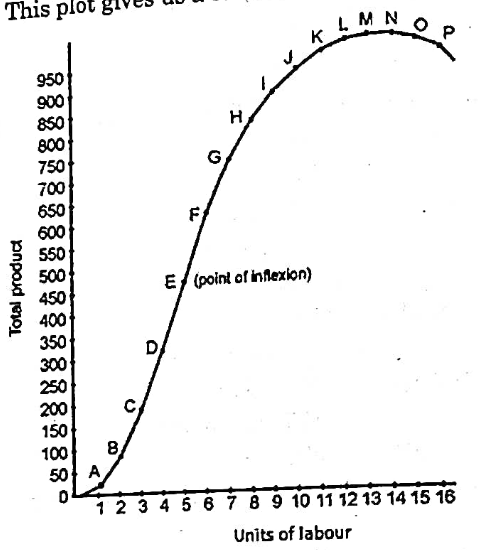

Fig. 7.1

Note that the total product curve is of the shape of an elongated or a stretched-out S. Also note that capital, the fixed input, is not shown in the diagram for the reason that in the short run, with which we are concerned here, total product increases because of the increase in variable input only.

As mentioned above, total product curve is obtained by plotting the total product against the number of units of variable product. Specifically, when L is 0 unit, that is, labour is not employed yet, TP is 0. By plotting this combination we get the point O in the diagram. When L is 1 unit, TP is 20 units. The plot of this combination of L and TP gives us point A and so on. By joining all such points we get total product curve.

Total product curve starts at 0 which means nothing will be produced if labour is not employed. Total product curve is rising continuously till the employment of 13th unit of labour. It is flat for a while and then it starts falling.

Total product curve can be divided into three parts with each part having different characteristics. These parts are

Part 1 : from point O to point E

Part 2 : from point E to point M

Part 3 : from point M onwards.

In the first portion, TP curve is getting steeper. Remember that technically steepness of a curve is called its slope. Mathematically it is equal to the change in the variable measured on vertical axis divided by the change in the variable measured on horizontal axis. In our case, slope of the TP curve is equal to $\Delta TP/\Delta L$. Note that this expression is one way of defining the marginal product. It means slope of total product curve also gives the marginal product of variable input. Slope of a curve can also be visualised as the rate of change in the variable on vertical axis with respect to the variable on horizontal axis, that is, by how much the variable on vertical axis changes when the variable on horizontal axis changes by a small amount. In our case, 'change by a small amount' is one unit change, that is, by how much TP changes when L is increased by 1 unit.

Thus, in the range O to E, the slope of TP curve is increasing or TP is increasing at an increasing rate. Note that 1st unit of labour adds 20 units to TP, 2nd unit adds 60 units, 3rd unit adds 95 units, 4th unit adds 125 units and 5th unit of labour adds 140 units to TP. This means additions made to total product by successive units of labour, in the range O to E, are increasing. This is the meaning of the phrase 'TP is increasing at an increasing rate.' This also means that the slope of TP curve is increasing.

Now, it will be easy to see that in the range E to M, additions made to TP by the successive units of labour are getting smaller and smaller. 6th unit of labour adds 135 units to TP which is less than the addition of 140 units, made by 5th unit of labour. Similarly, addition made by 7th unit (110 units) is less than that made by 6th unit of labour and so on up to 12th unit of labour. Thus, in the range E to M, TP is increasing but at a diminishing or decreasing rate. This also means that slope of TP curve is decreasing in this range.

At point M, the TP curve is plst. In other words the slope of the curve is zero; there is no change in TP when quantity of labour is increased from 13 to 14 units. The addition made by 14th unit is zero.

Beyond point M, TP curve is falling, that is, it has a negative slope: as more units of labour are added to the fixed quantity of capital, TP instead of rising is falling—successive units of labour are making a negative contribution to the production of the commodity.

In short, in the 1st portion, TP is increasing at an increasing rate, in 2nd portion TP increases at diminishing rate, in 3rd portion it does not change at all and finally in 4th portion, TP decreases.

Point E has a special feature. Before this point, TP is increasing at an increasing rate and past it TP increases at a diminishing rate. So at point F the rate of increase in TP changes (from increasing rate to decreasing rate). This point is called the point of inflexion or inflection. Thus, the point of inflexion is the point at which TP stops increasing at an increasing rate and starts increasing at a decreasing rate.

Remember that the total product curve is often referred to as (short run) production function because it gives graphically the functional relationship between output and inputs in the short run. (The diagram shows only one input- labour- but the other input- capital- is also there implicitly if not explicitly because, unaided, labour can do nothing).

Average Product Curve

Average Product Curve is obtained by plotting average product against the units employed of variable input. Average product of one unit of labour is 20. Plotting this combination of L = 1 and AP = 20 gives point A. The average product of two units of labour is 40. This combination of AP = 40 and L = 2 gives point B. Other points are obtained in the similar way. By joining these points we get Average Product Curve.

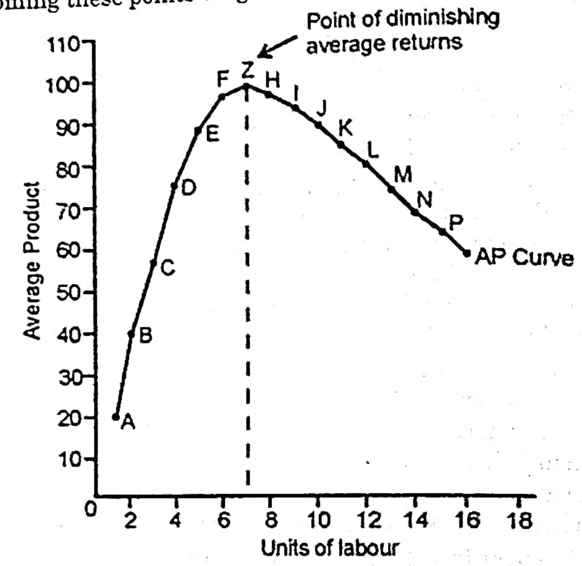

Fig. 7.2

Point Z is a special point. It is special in the sense that it gives the peak or highest point of the curve.

Note that Average Product Curve has two portions. One, from point A to point Z, and two, beyond point Z. In the 1st portion Average Product Curve has a positive slope: it is rising. It means Average Product is increasing as quantity employed of labour is increased. The curve reaches its highest point at Z. Beyond point Z, average product curve has a negative slope: it is decreasing. It means in this part average product is falling. The point where average product reaches its maximum, is called point of diminishing average returns because past this point average product falls continuously.

Note also that the shape of the average product curve is like that of an inverted U.

Marginal Product Curve

Marginal product curve of a variable input is obtained by plotting MP against the units of variable input used. Marginal product of 1st unit of labour is 20. Plotting this combination of MP = 20 and L = 1 gives point B. Marginal product of 2nd unit of labour is 60. The plot of this combination of MP = 60 and L = 2 gives point C. Other points are obtained in the similar way. By joining these points we get marginal product curve.

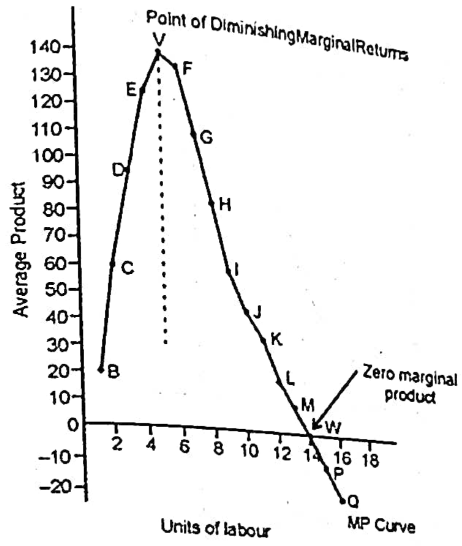

Fig. 7.3

Points V and W are special points. Point V is special in the sense that it gives the peak or highest point of the curve and point W is special because it represents the zero MP of the variable input. Zero MP means that the corresponding unit of variable input is doing nothing.

Note that Marginal product curve has three portions : first from point B to point V, the second from point V to point W and the third beyond point W. In the 1st portion, marginal product curve has a positive slope: it is rising. It means marginal product is increasing as quantity employed of labour is increased. The curve reaches its highest point at V. Beyond point V, marginal product curve has a negative slope: it is decreasing. It means in this portion marginal product is falling. But there are two parts in the decreasing portion of the curve: one from point V to point W and the other beyond point W. between V and W, marginal product though falling is positive- the units of variable input employed in this range are making some positive contribution to the total product but the rate of contribution is decreasing. Beyond point W, the units of variable input are making negative contribution to the production of the good- they obstruct the work done by others.

The point where Marginal Product reaches its maximum, is called point of diminishing marginal returns because past this point marginal product falls continuously.

6. Give the relationship as it exists among different production concepts.

Ans. There are some interesting and in-sighting relations among the three production concepts, the knowledge of which is very helpful in understanding the law of variable proportions. Some of these relations are listed below:

Relationship between TP and MP

MP is the rate of change of TP with respect to variable input, that is, MP is the change in TP resulting from the change in units employed of variable input (the change must be a

small change, usually one unit). Hence, it is the behaviour of MP which determines the behaviour of TP or the shape of total product curve.

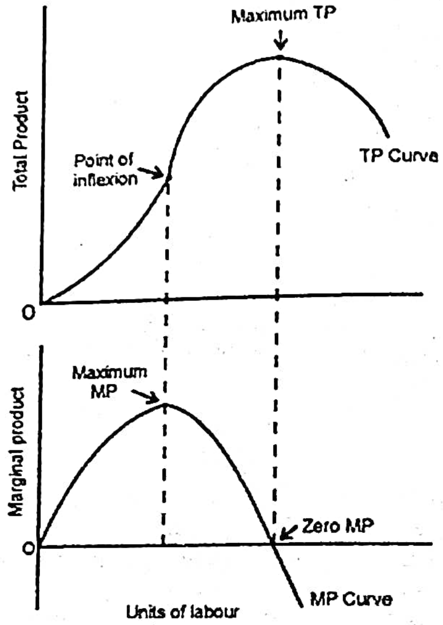

Fig. 7.4

- When MP is increasing, TP is increasing at an increasing rate. It also means the slope of total product curve is increasing because MP gives the slope of TP curve.
- When MP decreases, total product increases but at a decreasing rate. This also means the slope of total product curve is decreasing.
- When MP reaches its maximum, the rate of change of TP changes; before this maximum, TP increases at an increasing rate and after it TP increases at a decreasing rate. In terms of slope of TP curve, before the maximum level of MP, the slope of TP curve is increasing and after it the slope is decreasing. Thus, the maximum point of MP corresponds to the point of inflexion of TP curve.
- When MP becomes negative, TP falls. The reason is that the MP is the addition in TP, so when MP is negative it means there are negative additions to the TP which simply means TP is decreasing. In terms of slope of TP curve, negative MP means the negative slope of TP curve. Hence, negative MP corresponds to the falling portion of TP curve.
- When MP becomes zero, TP reaches its maximum. The logic is that zero MP means zero addition to TP, hence no change in TP. Before its zero level, although very small and decreasing, MP is positive. So, TP is increasing, though at a very small rate. After zero level, MP becomes negative. So, TP falls. Thus, it is at zero level of MP that TP is maximum. In terms of slope, zero MP means zero slope of TP curve

or total product become flat or parallel to horizontal axis: there is no change in TP as amount of variable input is increased.

Relation between MP and AP

Remember that the relationship between marginal and average values is a mathematical one. Hence it is not restricted to production concepts only. Rather, it applies to all other marginal and average concepts. Even the scope of this relationship goes beyond the field of economics.

This general relationship between average and marginal valves is as follows:

- When marginal value is higher than average value, average value rises.
- When marginal value is less than average value, average value falls.
- When marginal value is equal to average value, average value does not change.

A typical and common example, outside the field of economics, is that of the runs scored by a cricketer. Suppose a cricketer has played 5 innings. His score in these innings is 50, 42, 63, 30 and 25. His average score is $\frac{(50+42+63+30+75)}{5} = \frac{260}{5} = 52$. His latest score, 75 runs, is his marginal score. Suppose in his next inning he scores 64 runs. His marginal score fell from 75 to 64 runs. However his average is now $\frac{324}{6} = 54$. It is still rising because his marginal score of 64 is above his previous average of 52. Suppose again, in his next inning he scores 33, his average falls: it is now 51 because his marginal score of 33 is less than his previous average of 54. Finally suppose, in his next match the cricketer scores 51 runs which are exactly equal to his average score. His average score does not change. It remains constant at 51 runs.

Now, let us see the specific relation between average product and marginal products. This relationship between the two is listed below and shown in the following diagram.

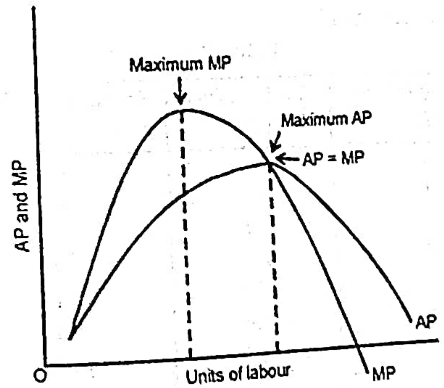

Fig. 7.5

- Both AP and MP first increase, reach a maximum and then decrease.

- Maximum level of MP is reached at a lower level of the employment of variable input than the maximum level of AP.
- The rates of increase as well as decrease in MP are greater than their AP counterparts.
- When MP is greater than AP, AP rises; it does not matter whether MP itself is increasing or decreasing.
- When MP is less than AP, AP falls.
- Since AP is rises or falls depending on whether MP is greater than or less than AP, it follows that MP must equal AP at the maximum level of AP.
- MP can be positive, zero or negative. But AP is positive; it can neither be zero nor negative.

7. Explain the law of variable proportions with the help of a table and a diagram.

Ans. The law of variable proportions is a short-run concept of production function. It gives us the variations in output that result from employing different quantities of a variable input in combination with given amounts of fixed inputs. Also remember that the law states that **if, in a given state of technology, increasing quantities of a variable input are applied to a given quantity of a fixed input, the marginal product, and the average product, of the variable input will sooner or later decrease.**

Table 7.2, and figures 7.5 and 7.6 give the tabular and graphic presentation of the law. (For our convenience, we re-present the table 7.1 as table 7.2 below without any change.)

The law, which gives the behaviour of output when the varying quantity of one factor is combined with the fixed quantity of other factor, is divided into three stages. But there are two approaches to define and distinguish each of the three stages. One approach takes in to consideration the behaviour of marginal as well as average products, and the other approach considers the behaviour of marginal product only. But this has to be noted that the first approach is more common.

Table 7.1

Units of capital (*i*)	*Units of labour* (*ii*)	*Total product* (*iii*)	*Average product* (*iv*)	*Marginal product* (*v*)
5	0	0	-	-
5	1	20	20	20
5	2	80	40	60
5	3	175	58	95
5	4	300	75	125
5	5	440	88	140
5	6	575	95.8	135
5	7	685	97.8	110
5	8	770	96.2	85
5	9	830	92	60
5	10	875	87.5	45
5	11	910	82.7	35
5	12	930	77.5	20
5	13	940	72.3	10
5	14	940	67	0
5	15	930	62	-10
5	16	910	56.8	-20

To distinguish the two approaches, the three stages are named differently under each approach. Under the first approach, that is, when average as well as marginal products are taken in to consideration, the three stages are named as:

1. Stage of increasing returns
2. Stage of diminishing returns, and
3. Stage of negative returns

Under the second approach, that is, when the behaviour of only the marginal product is considered, the three stages of the law are named as:

1. Stage of increasing marginal returns
2. Stage of diminishing marginal returns, and
3. Stage of negative marginal returns

Let's first take the three stages as they are defined under first approach, that is, when the behaviour of both marginal and average products is taken into consideration. Graphically, these are shown as in the following diagram.

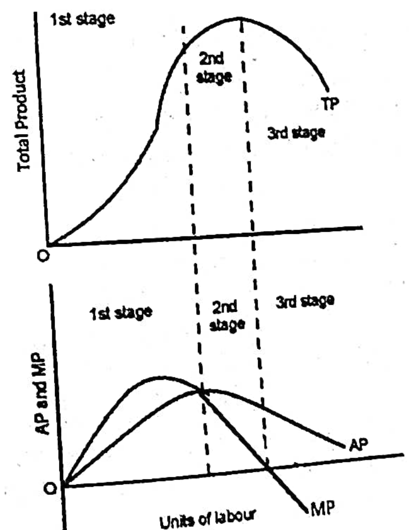

Fig. 7.6

1. **Stage of increasing returns :** This stage is identified by rising average product. In this stage, total product increases, first at increasing rate and then at diminishing rate. This stage begins at the employment of first unit of variable input and ends at the level at which average product is maximum. Since the average product is rising throughout in this stage, it implies that marginal product is greater than average product and, as explained elsewhere, it is this fact (marginal product being greater than average product) that average product is rising. Note that in this stage marginal product first increases and then decreases. But this is not important. What is important is that marginal product is greater than average product. Also, it is the increasing and decreasing stages of marginal product which lead total product to increase first at an increasing rate and then at decreasing rate. This stage is called stage of increasing returns because of increasing trend of average product of variable input.
2. **Stage of Decreasing Returns :** This stage is characterised by falling average product. In this stage, total product continues to increase but at diminishing rate and it is at the end of this stage that total product reaches its highest level. This stage begins at the level of employment of variable input at which average product is maximum and ends at the level at which marginal product is zero (and hence total product is maximum). Another feature of this stage is that marginal product is throughout less than average product. This is the reason for falling trend of average product. Also, it is the continuous fall in marginal product which results in total product increasing at diminishing rate. This stage is called stage of decreasing or diminishing returns because of the falling trend in both marginal as well as average products.
3. **Stage of negative Returns :** This stage is identified by negative marginal product. In this stage, total product falls and falls continuously. It does not rise again. This stage has no relation with the behaviour of average product. This stage begins at the level of employment of variable input at which its marginal product is zero. But this stage has no end: production will continue endlessly in this stage if the producer continues his production process. As regards average product, it continues to fall endlessly but it will remain positive. Total product falls because of negative marginal product. This stage is called stage of negative returns because of negative trend in marginal product.

Now we shall take the definition of these three stages on the basis of the behaviour of marginal product only. Graphically, these are shown as in the following diagram.

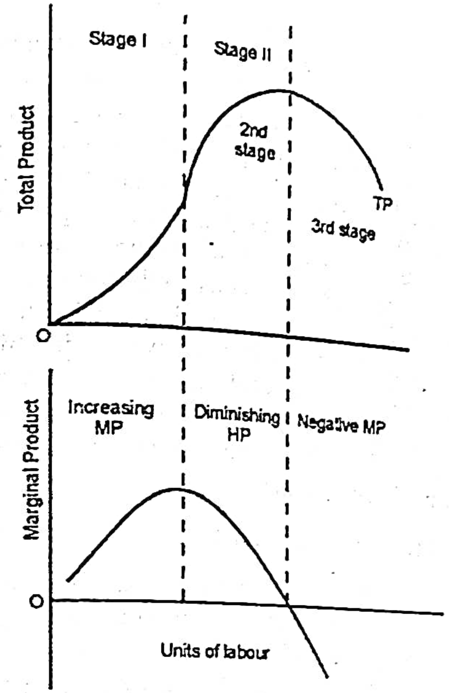

Fig. 7.7

1. **Stage of Increasing Marginal Returns :** This stage is identified by rising marginal product. Because of rising marginal product, total product increases at increasing rate. This stage begins at the employment of first unit of variable input and ends at the level at which marginal product is maximum. It is to be noted that maximum level of marginal product is also the point at which it stops increasing.
2. **Stage of Diminishing Marginal Product :** This stage is identified by failing marginal product. Because of falling marginal product, total product increases at decreasing rate. This stage begins at the level of employment of variable input at which marginal product is maximum and ends at the level at which it is zero.
3. **Stage of Negative Marginal Returns :** This stage is identified by negative marginal product. Because of negative marginal product, total product decreases. This stage begins at the level of employment of variable input at which marginal product is zero and continues endlessly.

8. Give the reasons for the operation of law of variable proportions.

Ans. How output varies in short-run as a producer varies quantity of some input in relation to a fixed quantity of other input is given by the law of variable proportions. But why does the output behave as described by the law. In other words, what are the reasons for the operation of the law of variable proportions?

To be very precise, it is the behaviour of marginal product which makes total product to behave as described by the law. In other words, the distinguishing feature of the three stages of the law is the way in which marginal product behaves.

Take first the approach of distinguishing the three stages of the law based on marginal product only. In this approach it is straight forward that when marginal product is rising, we have increasing marginal returns, when marginal product is falling but remains positive, we have diminishing marginal returns, and when marginal product is negative, we have negative marginal returns. Under average product-marginal product approach, when marginal product is greater than average product so that average product rises, we have increasing returns to a factor. Next, when marginal product is less than average product but is positive, so that average product falls, we have diminishing returns to a factor. Finally, when marginal product becomes negative, we have negative returns to a factor.

All this implies that to explore the reasons for the operation of law of variable proportions or returns to a factor, we have to find the reasons which make marginal product to behave as described above. Let's find these reasons:

Reasons for Increasing Returns to a Factor :

Loosely speaking, increasing returns to a factor occur when marginal product is rising and is greater than average product so that total product is increasing at an increasing rate. The reasons for rising marginal product are:

1. **Fuller and Efficient Utilization of Fixed Factor :** In general, fixed factors are available in large sizes and are indivisible so that they cannot be taken in quantities which suit a smaller quantity of the variable factor. Thus in the beginning, the quantity of fixed factor is abundant relative to the quantity of variable input, that is, too much of fixed factor is working with too little of variable input, so that there is incomplete and inefficient utilization of fixed factor. When the quantity used of variable input is increased, fuller, intensive and efficient utilization of fixed factor takes place leading to higher output per unit of input.
2. **Division of Labour and Specialization :** As more units of the variable input are used, scope for the introduction of division of labour and specialization is created. The introduction of division of labour and specialization increases efficiency of variable input leading to higher output.

Reasons for Diminishing Returns to a Factor:

Diminishing returns to a factor occur when marginal product is falling and is below average product so that total product increases at a diminishing rate. The reasons for falling marginal product are:

1. **Inadequacy of Fixed Factor in Relation to Variable Factor :** There is a limit in the use of the amount of variable input which ensures the complete, optimum and efficient utilization of indivisible fixed factor used with it. Once this point is reached, further increase in variable factor results in the scarcity of fixed factor relative to the quantity used of variable factor. In other words, the fixed factor becomes inadequate in relation to the quantity of variable factor: quantity of fixed factor is insufficient to ensure optimum utilization of variable factor. The inefficient utilization of variable input causes marginal and average product to decline.
2. **Imperfect Substitutability of Factors :** Imperfect substitutability of factors is another cause for diminishing returns. Substitutability of factors means the ability to replace one factor with another. Perfect substitutability of factors means the ability a factor completely replaces the other and imperfect substitutability is a situation where in in which one factor partially replaces the other. In real life, factors are imperfect substitutes. Had the factors been perfect substitutes, the scarcity of fixed factors in the second stage would not have mattered. Producer could add more and more of variable input. This would make up for the scarcity of fixed factor and returns would remain constant. But factors being imperfect substitutes, addition of more units of variable input does not make up for the scarcity of fixed factor. The result is fall in the marginal and average product.

Reasons for Negative Returns to a Factor

Negative returns to a factor occur when marginal product becomes negative and total product falls. This happens for the following reason:

Too Much Quantity of Variable Input : The phenomenon of negative returns is due to the fact that the units of variable factor become too excessive relative to the fixed factor so that they get in each other's way with the result that the total output falls instead of rising. In other words, due to excessive quantity of variable factor, there would be congestion, overcrowding, lack of coordination, mismanagement and lack of supervision which ultimately result in decrease in production

9. In which of the three stages of law of variable is a producer going to operate?

The law of variable proportions has three stages; stage of increasing, diminishing and negative returns. The question arises, in which stage a producer will undertake the process of production. The common sense may say that this will be done in the stage of increasing returns because in this stage the producer gets output at an increasing rate: the average as well as the marginal product of the variable input are increasing and in other two stages the returns are either diminishing or negative.

But the common sense is not always the right sense. The above issue is one such case. When we seek the answer to the above question from economic theory, the answer we get is that a rational producer will produce in stage II, that is, the stage of diminishing returns. The explanation for this is as follows;

A rational producer is one who weighs the costs and benefits of alternative courses of action and then seeks to maximise its net benefit. When a producer produces in the stage of increasing returns, he is not making the best possible use of the fixed factor and there still remains the scope for increasing the production by increasing the quantity of variable factor. Thus, a rational producer will not stop in the stage of increasing returns but will expand his production capacity further.

Stage of negative returns is out rightly out of question. Why should a producer, when he acts rationally, produce in this stage when the contribution of variable input is negative. He will never do so. Even, a rational producer will never think of entering this stage.

Thus, we find that a rational producer will produce only in the stage of diminishing returns where, although, the returns of variable input are decreasing but the producer is making fuller utilization of factor inputs employed by him.

Having deduced that the stage of diminishing returns is the operational stage for a rational producer, another question is that at what point of the stage of diminishing returns a producer will produce. The answer to this question depends on the cost of production which in turn depends on the prices of factor inputs. (Because of limited scope of this book, we will not take this issue any further)

10. Explain the law of constant proportions with the help of relevant tables and diagrams.

Ans. The law of constant proportions or returns to scale gives the general tendency of change, in long-run, in the output of a good as the quantity of all inputs is changed in such a way that the input ratio remains constant. Mathematically,

$$\text{Returns to scale} = \frac{\text{percentage change in output}}{\text{percentage change in inputs}}$$

Statement of the law : The law states that in the long-run, as a firm increases the quantities of all inputs employed in such a way that the factor ratio remains constant at all levels of output, the output may rise either more than proportionately, less than proportionately or in exactly same proportion as the change in inputs.

As is evident from the above statement of the law of variable proportions, this is a set of three cases. These are;

1. When output increases more than proportionately as the increase in inputs
2. When output increases less than proportionately as the increase in inputs
3. When output increases in exactly the same proportion as the increase in inputs

The first case is called law of increasing returns to scale, the second case is called law of diminishing returns to scale and the third case is called law of constant returns to scale.

Let us study these three laws in some detail.

1. Increasing Returns to Scale

Increasing returns to scale is a situation in which a given proportionate increase in all inputs leads to an increase in output which is more than the proportionate increase in inputs.

Following table explains this situation in detail.

Capital (K)	*ΔK*	*%ΔK*	*Labour (L)*	*ΔL*	*%ΔL*	*L/K*	*Output (Q)*	*ΔQ*	*%ΔQ*
(I)	*(II)*	*(III)*	*(IV)*	*(V)*	*(VI)*	*(VII)*	*(VIII)*	*(IX)*	*(X)*
2			6			3	10		
4	2	100	12	6	100	3	22	12	120
6	2	50	18	6	50	3	35	13	59
10	4	66.6	30	12	66.6	3	50	15	42.8
14	4	40	42	12	40	3	75	25	50
20	6	42.8	60	18	48.8	3	115	40	53.3

We have assumed that there are two inputs capital (K) and labour (L). Since we are in the long run, we can change the quantities of both the inputs. The quantities are increased in such a way that the ratio of inputs remains fixed. This is shown by column (VII). This column gives the ratio of labour to capital (K/L) which remains constant (3) at all levels of output. Columns (I) and (IV) give the quantities used of capital and labour inputs respectively. Columns (II) and (V) give absolute changes in the two inputs. Columns (III) and (VI) give the percentage changes in labour and capital inputs. Column (VIII) gives quantity of output produced at various levels of inputs. Column (IX) gives absolute changes in output and column (X) gives percentage changes in output. In Columns (III) and (VI) we find that the percentage change in the two inputs is same. This is necessary if the ratio of the inputs is to be kept constant at all levels of output.

Comparing the percentage changes in inputs with the percentage changes n output, we find that the percentage changes in output are greater than the percentage changes in input. For example, when inputs are doubled, that is, increased by 100% - capital from 2 units to 4 units and labour from 6 units to 12 units- output increases by more than double from 10 to 22 units, that is, by 120%. Next, when inputs are increased by half or 50% - capital from 4 to 6 and labour from 12 to 18- output increases by more than half or by 59% and so on. This is what is implied by law of increasing returns to scale: an increase in all inputs, at some constant ratio, leads to a more than proportionate increase in the level of output.

Percentage change in output > percentage change in inputs

2. Diminishing returns to Scale

Diminishing returns to scale is a situation in which a given proportionate increase in all inputs leads to an increase in output which is less than the proportionate increase in inputs.

Following table explains this situation in detail.

Capital (K)	ΔK	$\%\Delta K$	Labour (L)	ΔL	$\%\Delta L$	L/K	Output (Q)	ΔQ	$\%\Delta Q$
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
2			6			3	15		
4	2	100	12	6	100	3	25	10	66.6
6	2	50	18	6	50	3	37	12	48
10	4	66.6	30	12	66.6	3	60	23	62
14	4	40	42	12	40	3	80	20	33.3
20	6	42.8	60	18	48.8	3	112	32	40

Comparing the percentage changes in inputs with the percentage changes n output, we find that the percentage changes in output are smaller than the percentage changes in input. For example when inputs are doubled, that is, increased by 100% - capital from 2 units to 4 units and labour from 6 units to 12 units- output less than double from 15 to 25 units, that is, by 66.6%. Next, when inputs are increased by half or 50% -capital from 4 to 6 and labour from 12 to 18- output increases by less than half or by 48% and so on. This is what is implied by law of diminishing returns to scale: an increase in all inputs, at some constant ratio, leads to a less than proportionate increase in the level of output.

Percentage change in output < percentage change in inputs

3. Constant Returns to Scale

Constant returns to scale is a situation in which a given proportionate increase in all inputs leads to an increase in output which is proportional to the increase in inputs.

Following table explains this situation in detail.

Capital (K)	ΔK	$\%\Delta K$	Labour (L)	ΔL	$\%\Delta L$	L/K	Output (Q)	ΔQ	$\%\Delta Q$
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
2			6			3	10		
4	2	100	12	6	100	3	20	10	100
6	2	50	18	6	50	3	30	10	50
10	4	66.6	30	12	66.6	3	50	20	66.6
14	4	40	42	12	40	3	70	20	40
20	6	42.8	60	18	48.8	3	100	30	42.8

Comparing the percentage changes in inputs with the percentage changes n output, we find that the percentage changes in output are exactly equal to the percentage changes in input. For example when inputs are doubled, that is, increased by 100% – capital from 2 units to 4 units and labour from 6 units to 12 units- output also doubles from 10 to 20 units, that is, by exactly 100%. Next, when inputs are increased by half or 50% – capital from 4 to 6 and labour from 12 to 18- output increases by the same proportion from 20 to 30 units

and so on. This is what is implied by the law of constant returns to scale: an increase in all inputs, at some constant ratio, leads to a proportional increase in the level of output.

Percentage change in output = percentage change in inputs

Production functions which exhibit constant returns to scale possess convenient mathematical properties making it very useful for theoretical analysis. Mathematically, the production function exhibiting constant returns to scale is called linearly homogeneous production function or production function of degree one.

11. What are the causes of emergence of different cases of returns to scale?

Ans. To be very brief and exact, different cases of returns to scale are the result of the emergence of economies and diseconomies of scale. Economies of scale are the benefits of higher productivity of factor inputs which accrue to a firm because of production of the good it produces being undertaken on a larger scale or size. Diseconomies of scale are the costs or disadvantages of reduced productivity of factor inputs resulting in higher cost per unit of output which accrue to a firm because of production of the good it produces being undertaken on a very large scale or size.

When a firm experiences economies of scale, productivity of inputs increases and we have increasing returns to scale, and when it experiences diseconomies of scale, productivity of inputs decreases and we have decreasing returns to scale. Also, when diseconomies of scale start emerging, the expanding firm may just succeed in balancing diseconomies of scale with economies of scale, so that constant returns to scale prevail. There is also the possibility that a very small rate of output may enable firms to capture all the economies of scale. Diseconomies may not be faced until the rate of output increases to very high levels. The time interval in which there are no economies or diseconomies of scale may also enable a firm to experience constant returns to scale.

Economies of scale can be real or pecuniary. Real economies of scale are the benefits of large scale production associated with the reduction in the use of physical quantity of inputs per unit of output. These benefits accrue to a firm from the factors related to labour, machinery, and management etc. The benefits which are related to labour are called labour economies, those related to machinery or for that matter capital are called technical economies, and those related to management are called managerial economies of scale.

Pecuniary economies of scale are the benefits of large scale production associated with the payment of lower prices on account of bulk-buying of raw materials, increase in bargaining power etc. These economies accrue to a firm from the factors related to selling and marketing, transportation, finance etc. To distinguish them from pecuniary economies, real economies are often referred to as non-pecuniary economies of scale.

Each of the real and pecuniary economies of scale can occur to a firm from the factors which are either internal or external to it. Scale economies which accrue to a firm because of factors which are internal to it are called internal economies of scale and those which

accrue to it because of factors external to it are called external economies of scale. More precisely, internal economies of scale are those scale economies which accrue to a firm because of the increase in its own scale of production. External economies of scale, on the other hand, are those scale economies which accrue to a firm because of the expansion of the output of the entire industry in which the firm is producing. Internal economies are firm specific: they accrue to that firm only which produces at a larger scale. External economies are available to all firms irrespective of the fact that whether a firm expands its own output or not.

It must be now clear that we have four categories of economies of scale. These are:

1. Internal-real economies of scale
2. Internal-pecuniary economies of scale
3. External-real economies of scale, and
4. External-pecuniary economies of scale

Since we are analysing the laws of returns to scale or the long-run production function and since the production function describes the relationship between inputs and output expressed in physical terms, the economies which are relevant for this analysis are internal-real economies and diseconomies of scale. The other three categories are irrelevant for the moment.

CHAPTER 8

Isoquants

1. What are isoquants? What is an isoquant map?

Ans. An isoquant derived from two Greek words 'iso' meaning same and 'quant' meaning quantity, also called iso-product curve, equal product curve, a producer's indifference curve or equal quantity curve, is a line showing all the alternative combinations of two factors that can produce a given level of output. The concept can be explained as follows:

Imagine that a firm wants to produce a certain level of output: say, 5000 units per year. Let us assume that it estimates all the possible combinations of labour and capital that could produce that level of output. Some of these estimates are shown in the following table.

Combination	A	B	C	D	E
Units of labour	2	4	5	8	12
Units of capital	12	7	6	4	3

The graphic representation of these alternative combinations of inputs to produce a given level of output gives an isoquant.

The following figure shows the 5000 unit isoquant corresponding to the above table.

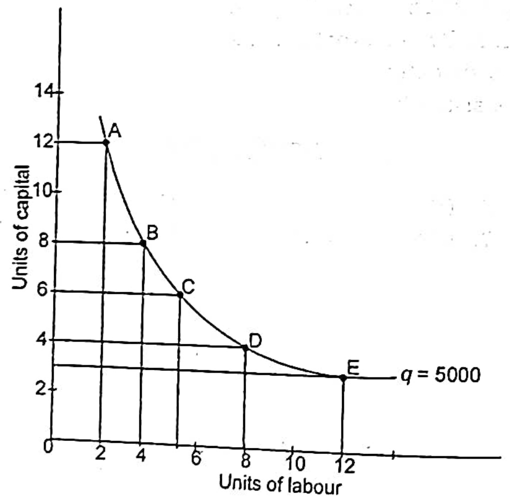

Fig. 8.1

Horizontal axis of the diagram shows units of labour employed and vertical axis shows units of capital. The curve labelled as Q = 5000, is an isoquant giving various combinations of labour and capital which can produce 5000 units of output.

It is to be noted that an isoquant shows only technically efficient combinations of the two inputs. That is, the curve shows the minimum level of each input necessary for given levels of output and other inputs. In other words, any point on an isoquant shows the minimum quantities of inputs needed to produce the given level of output.

It is also to be remembered that an isoquant shows the whole range of alternative ways of producing a given output. Thus, the above figure shows not only points A to E from the table, but all the intermediate points too.

Isoquant Map

Like indifference curves, a whole series of isoquants can be drawn, each one representing a different level of output. Illustrating a number of isoquants on the same diagram gives us an isoquant map, which represents a number of isoquants for a given producer with reference to two inputs. In other words, an isoquant map is a collection of isoquants that represent a given producer's entire options with reference to two inputs, with each isoquant corresponding to a different level of total physical product,

Following diagram shows three isoquants—q_1, q_2 and q_3, where q_1 represents a level of output which is less than that represented by q_2, and q_2 represents the level of output which is less than that represented by q_3. In short, the higher the output, the further out to the right will the isoquant be.

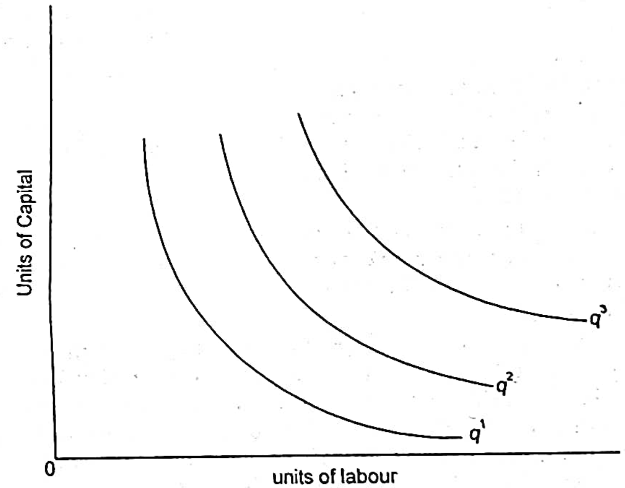

Fig. 8.2

Although only three isoquants have been drawn, many more could have been added. For example, many isoquants lie between q_1 and q_2.

2. Explain the principle of diminishing marginal rate of technical substitution.

Consider the following figure in which a producer is producing at point F of isoquant q_1 using, say 14 units of capital and 2 units of labour.

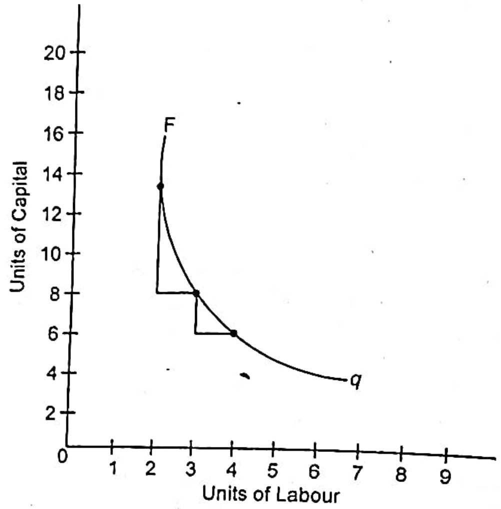

Fig. 8.3

If the producer wants to use an additional unit of labour, does he need to give up an additional unit of capital and still produce the same level of output as at point F. The answer is no. the reason is that at F the producer is using more capital relative to labour so that marginal product of labour is more than the marginal product of capital. If the producer gives up one unit of capital and adds one additional unit of labour, his gain in output due to the use of an additional unit of labour is more than his loss due to the reduction of capital input by one unit so that he is not producing the same level of output as at F. Rather his total product is more than at F.

It means at F the producer has to give up more than one unit of capital for an additional unit of labour if his total product has to remain same as at F. Let us suppose that the producer has to give up 6 units of capital for an additional unit of labour in order to produce the same level of output as at F.

The amount of one input, say capital, a producer is willing to give up to add an additional unit of another input, say labour, and still maintain the existing level of output, is called marginal rate of technical substitution or marginal rate of factor substitution of labour for capital. In other words, the marginal rate of technical substitution of labour for capital (MRTSL,K) tells us how many units of capital the firm can replace with an extra unit of labour while holding output constant. Mathematically, we can write

$$MRTS_{L,K} = \frac{\Delta K}{\Delta L}$$

This says that marginal rate of technical substitution of labour (the input measured on horizontal axis) for capital (the input measured on vertical axis) is the quantity change in capital (ΔK) necessary to keep output constant if labour changes by ΔL.

In our example, at point F marginal rate of technical substitution of capital for labour is 6 units of capital for one unit of labour.

$$\text{MRTS at F} = \frac{\Delta K}{\Delta L} = \frac{6}{1} = 6$$

At a different point than F, marginal rate of technical substitution between capital and labour will be different. In principle marginal rate of technical substitution for an input diminishes as a producer uses more of it. For example, at point G, the producer is using more units of labour and fewer units of capital than at F. Hence, he will not have to give up 6 units of capital for an additional unit of labour. At this point he would give up less than 6 units of capital for an additional unit of labour so as to keep the level of output unchanged. Let us say that at point G, marginal rate of technical substitution is 1, that is, 1 unit of capital for one unit of labour which is less than 6 units of capital for an additional labour at point F.

In short, the principle of diminishing marginal rate of technical substitution says that the amount of one input a producer has to give up when using an additional unit of another input so as to keep output constant, decreases as the producer employs more of the second input.

The principle of diminishing marginal rate of technical substitution can be explained in more detail with a table and a diagram as follows:

Factor Combinations	*Units of Labour*	*Units of Capital*	*MRTS of L for K* = $\frac{\Delta K}{\Delta L}$
A	1	12	—
B	2	8	4
C	3	5	3
D	4	3	2
E	5	2	1

The table shows that in order to produce a given level of output (not shown in the table) a hypothetical producer can use any of the five combination labour and capital-A to E. Starting from combination A, the producer can use 1 unit of labour with 12 units of capital. It can also use 2 units of labour with 8 units of capital, and so forth.

The thing to be noted is that in the beginning the producer is using a larger quantity of capital relative to labour. In this case marginal product of labour must be very large than that of capital. If the producer wants to increase the quantity of labour, he has to decrease the quantity used of capital if the level of output is to remain fixed. At this point the producer has to give up 4 units of capital in order to add an extra unit of labour so that output remains constant. It means at this time $MRTS_{L,K}\left(\frac{\Delta K}{\Delta L}\right)$ equals 4 : (4/1 = 4). The producer is using combination B : 2 units of labour and 8 units of capital.

If the producer wants to use another unit of labour, he has to give up some more units of capital if the level of output is to remain constant. At this point, compared to point A marginal product of labour will have decreased and that of capital will have increased. The reason being that compared to combination A, the producer is using more labour (2 compared to 1) and less capital (8 compared to 12). As a result, the producer needs not to give up 4 units of capital for 1 unit of labour. Rather he has to give up only 3 units of capital for 1 unit of labour. It means at this point $MRTS_{L,K}$ $\left(\frac{\Delta K}{\Delta L}\right)$ equals 3 : (3/1 = 3). The producer is using combination C: 3 units of labour and 5 units of capital.

Repeating the exercise we can see that as the producer increases the quantity of labour, he has to give up lesser and lesser quantity of capital to maintain the given level of output which means MRTS of labour for capital goes on decreasing because increase in quantity of labour decreases its marginal product and decrease in quantity of capital increases it marginal product. At D $MRTS_{L,K}$ is 2 and at E it is 1.

Same thing can be explained diagrammatically as follows.

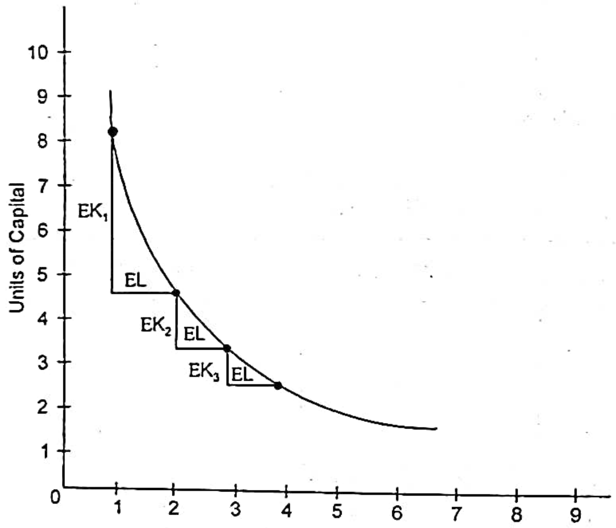

Fig. 8.4

Horizontal axis of the diagram shows labour units employed and vertical axis shows units of capital. The curve labelled as q is an isoquant giving various combinations of labour and capital that can produce a given level of output.

We start from point A. When the producer moves from point A to point B on the given isoquant, he gives up ΔK_1 of capital for ΔL of labour. Therefore, the $MRTS_{L,K}$ is $\frac{\Delta K_1}{\Delta L}$. As the producer slides down along the isoquant, keeping the length of ΔL fixed, the length of ΔK becomes shorter and shorter. It can be seen from the figure that ΔK_1 is longer than ΔK_2, ΔK_2 is longer than ΔK_3 and so on. It means as quantity of labour increases and that of capital decreases, the producer has to give up less and less of capital for a given increase in labour. In other words, the marginal rate of technical substitution of L for K diminishes as the quantity of labour increases and that of capital decreases.

3. Show that the marginal rate of technical substitution between two inputs equals the slope of an isoquant.

Ans. The fact that the marginal rate of technical substitution between two inputs equals the slope of an isoquant can be shown with the help of the following diagram:

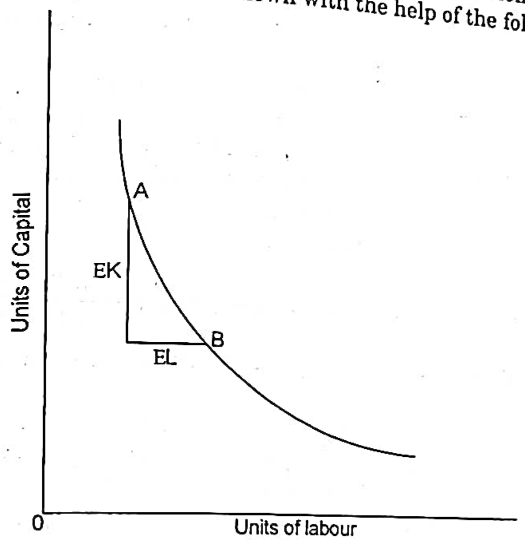

Fig. 8.5

The diagram gives an isoquant q which shows the alternative combinations of two inputs L and K that can produce a given level of output of a good. When the producer moves from any given combination, such as combination A, to another combination farther to the right on the curve, such as B, he must give up some of one input (K) in order to use more of the other input (L).

The slope of the isoquant at any point (*i.e.*, the slope of the line tangent to the curve at that point) is $\frac{\Delta K}{\Delta L}$ —the rate of change of K relative to the change of L.

But the expression $\frac{\Delta K}{\Delta L}$ is actually a measure of the marginal rate of technical substitution between the two inputs, K and L-the amount of K he would give up (ΔK) to gain additional L (ΔL).

In short, the slope of an isoquant at any point given by the slope of the line tangent to

the curve at that point gives the marginal rate of technical substitution between the two inputs represented by the isoquant.

However, because an isoquant slopes downwards from left to right, the slope of an isoquant is negative. It means, on a graph with L on the horizontal axis and K on the vertical axis, $MRS_{L,K}$ at any point is the *negative* of the slope of the isoquant at that point.

4. Give the relationship between marginal rate of technical substitution and marginal product.

Ans. The marginal product of any input is the increase in total product that a producer obtains from an additional unit of that input. Most inputs are assumed to exhibit diminishing marginal product: the more of the input a producer is already using, the lower the marginal product of that input.

The marginal rate of technical substitution between two inputs depends on their marginal products. For example, if the marginal product of input L is twice the marginal product of input K, then a producer would need 2 units of input K to compensate for losing 1 unit of input L, and the marginal rate of technical substitution equals 2. More generally, the marginal rate of technical substitution equals the marginal product of one input divided by the marginal product of the other input. That is,

$$\text{Marginal rate of tecnical substitution between L and K} = \frac{\text{MP of L}}{\text{MP of K}}$$

Where L is the input measured on horizontal axis and K is the input measured on vertical axis.

The same result can be shown in another way as follows:

As one moves down an isoquant, total output, by definition, will remain the same. Thus the loss in output due to less of one input, say capital, being used (*i.e.*, $MPP_K \times \Delta K$) must be exactly offset by the gain in output due to more labour being used (*i.e.* $MPP_L \times \Delta L$). Thus :

$$MPP_L \times \Delta L = MPP_K \times \Delta K$$

This equation can be rearranged as follows:

$$\frac{MPP_L}{MPP_K} = \frac{\Delta K}{\Delta L} = MRTS$$

Where MPP_L is the marginal physical product or simply marginal product of labour and MPP_K is the marginal physical product or simply marginal product of capital.

Thus the MRTS is equal to the inverse of the marginal productivity ratios of the two factors.

5. Describe the properties of ~~indifference~~ Isoquant ~~curves~~.

Ans. Like indifference curves, isoquants also exhibit some unique properties. Some of these properties include:

1. **Isoquants are downward sloping (from left to right).** We know that an isoquant gives alternative combinations of two inputs, say labour and capital,

which can produce a given level of output. Its downward sloping property follows from the fact that the marginal physical products of the factors are positive, that is, additions in the use of factors yield positive increments in output so that when the quantity of one factor, say labour, is increased, the quantity of other factor, that is, the capital, is to be decreased if the level of output is to remain unchanged along an isoquant as shown in the following figure with the help of direction of arrows.

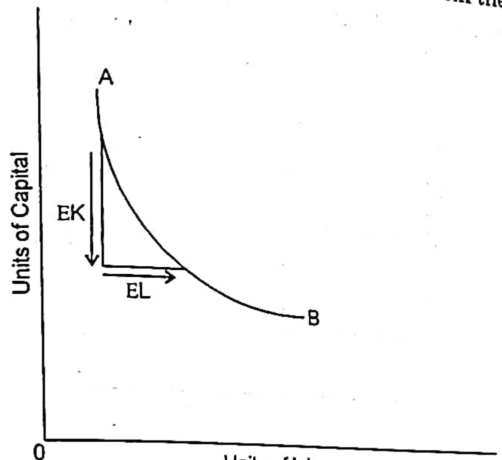

Fig. 8.6

If the isoquant is a horizontal straight line, as in the following figure, this would indicate that the marginal product of the factor measured on horizontal axis (usually labour) is zero, that is, the contribution of additional units of the factor is nil which means that the additions made in labour has no effect on level of output.

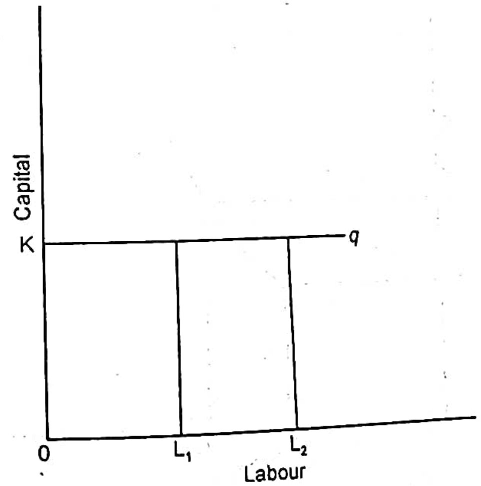

Fig. 8.7

The figure shows that whether the producer uses L_1 or L_2 units of labour with a given quantity of capital, output does not change which implies that additional labour makes no contribution in the production process.

Similarly, if the isoquant is a vertical straight line, as in the following figure, this would indicate that the marginal product of the factor measured on vertical axis (usually capital) is zero, that is, the contribution of additional units of the factor is nil which means that the additions made in capital has no effect on level of output.

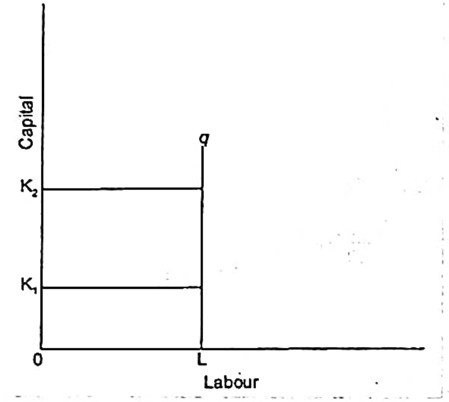

Fig. 8.8

The figure shows that whether the producer uses K_1 or K_2 units of capital with a given quantity of llabour, output does not change which implies that additional capital makes no contribution in the production process.

Finally, an upward sloping isoquant implies that the same output can be produced with the use of less of both the factors as shown in the following figure.

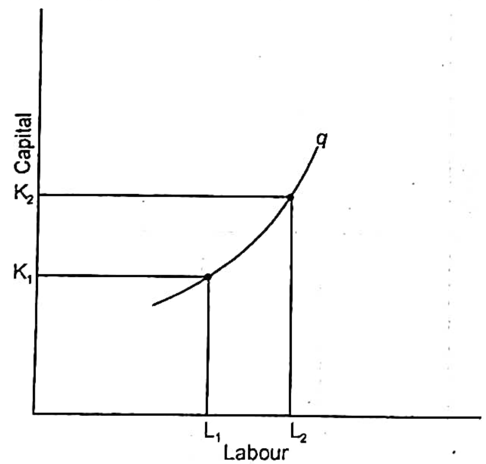

Fig. 8.9

This in turn implies that the marginal product of at least one factor is negative. But this is ruled out from the assumption of positive marginal product of factors which rules out the upward slope of an isoquant

2. **Isoquants that are farther from the origin, show the higher level of output.** The farther away from the origin an isoquant lies, the higher the total output is that it represents. This is because, the more inputs a firm uses, the more output it gets if it produces efficiently. For example, at point E in the following figure, the firm is producing 60 units of output with 2 workers and 7 units of capital. If the firm holds the number of units of capital constant and adds one more worker, it produces at point F. Point F must be on an isoquant with a higher level of output-here, 80 units-if the firm is producing efficiently and not wasting the extra labour.

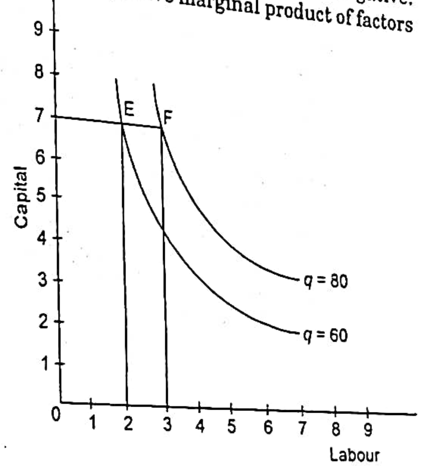

Fig. 8.10

3. **Isoquants do not cross (intersect).** We know that an isoquant shows only the technically efficient combinations of the two inputs. That is, the curve shows the minimum level of each input necessary for the production of given level of output. Efficiency requires that isoquants do not cross or intersect because such intersection is inconsistent with the efficiency requirement of an isoquant. Suppose that an isoquant representing 15 units of output (q = 15) intersects with an isoquant representing 30 units of output (q = 30) as shown in the following figure.

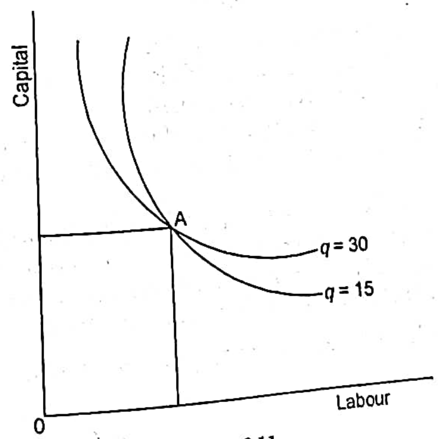

Fig. 8.11

The intersection of these two isoquants means that the factor combination which can produce 15 units of output can also produce 30 units of output.

If the firm produces 15 units of output with input combination with which it could produce 30 units of output, it must be producing inefficiently. Thus, the input combination which can produce 30 units of output should not lie on the isoquant representing 15 units of output which should include only efficient input combinations. In short, efficiency requires that isoquants do not cross.

4. **Isoquants are bowed inward (convex to the origin).** The principle of diminishing marginal rate of technical substitution which says that the amount of one input a producer has to give up to use an additional unit of another input, and maintain equal total output, decreases as the producer adds more of the input for which trade off is to be made, has an important implication for the shape of an isoquant: as we move down and to the right along an isoquant, we continue to substitute labour for capital as shown in the following figure.

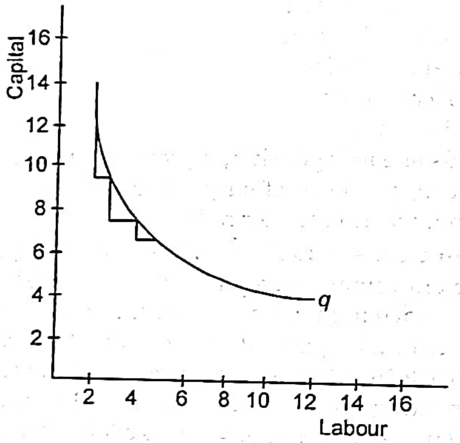

Fig. 8.12

As the firm produces the same quantity of output using more labour, the amount of capital it needs to give up for using additional units of labour decreases because of diminishing returns. This gives rise to the diminishing marginal rate of technical substitution. The geometric implication of this phenomenon is that as we move down and right along an isoquant, it becomes flatter making an isoquant convex to the origin.

If an isoquant is concave to the origin, it will imply that marginal rate of substitution of an input increases as more of it is obtained which is possible under increasing returns to a factor only. But it is the diminishing returns to a factor which is in operation because of which isoquants cannot be concave.

6. *Write a note on L-shaped and straight isoquants.*

Ans. An isoquant is a graphic exhibition of the alternative combinations of two inputs available to a producer for producing a given level of output. Generally, isoquant are convex

to the origin indicating the principle of diminishing marginal rate of technical substitution. Marginal rate of technical substitution, in turn, is a reflection of the substitutability between two inputs. When the inputs are easy to substitute for each other, the isoquants are less bowed or flatter; when the inputs are hard to substitute, the isoquant are very bowed or are relatively convex.

Take the example of a car manufacturing company. It can use either robots (capital) or humans (labour) to perform a given task, say painting the car. Suppose that a robot can do the work twice as quickly as humans. Here, the firm can always substitute one robot for two workers or vice versa, regardless of its current number of robots or workers. The marginal rate of technical substitution of capital for labour $\left(\frac{\Delta K}{\Delta L}\right)$ of the producer between robots and humans would be a fixed number — $\frac{1}{2} = 0.5$.

We can represent the alternative combinations of the two inputs in this production process with the help of an isoquant as in the following diagram.

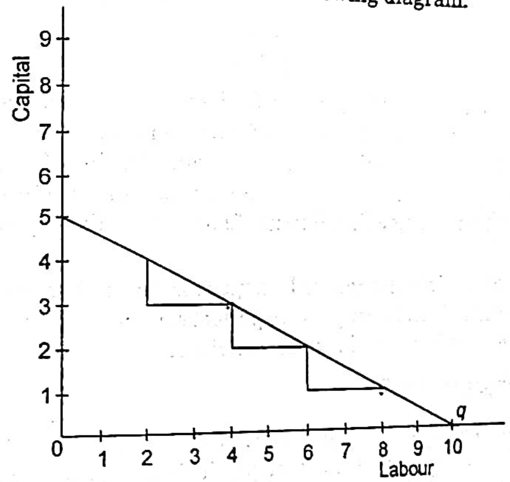

Fig. 8.13

Because the marginal rate of technical substitution is constant, the isoquants are straight lines, and the production functions they represent are called straight line production functions.

In this extreme case of straight isoquants, we say that the two inputs are perfect substitutes. Thus, two inputs are perfect substitutes in a production process when a producer can substitute them at a constant rate.

Now imagine a transport company which plies buses for passengers. To run a bus (capital), the company needs two persons (labour) -a driver and a conductor. It is useless for this company to purchase a new bus without hiring a new driver and a new conductor. Hiring a new driver and a new conductor without purchasing a new bus is useless as well.

We can represent the alternative combinations of one bus and two persons for the company with isoquants as in the following diagram.

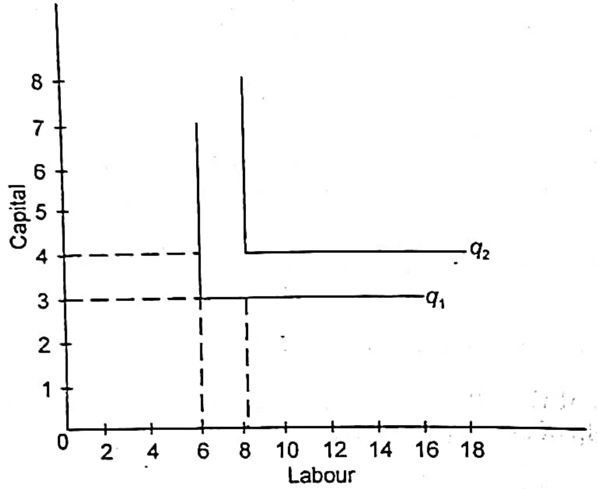

Fig. 8.14

In this case, a combination of 3 buses and 6 labourers (3 drivers + 3 conductors) is just as good as a combination with 3 buses and 8 labourers. It is also just as good as the combination with 4 buses and 6 labourers. The isoquants, therefore, are right angles or L-shaped. L-shaped isoquants are also called Leonitief isoquants and the production functions they represent are called Leonitief production functions or fixed proportions production functions.

In this extreme case of right-angle or L shaped isoquants, we say that the two inputs are perfect complements. Thus, two inputs are perfect complements in a production process when they need to be used in some fixed proportions only.

7. What is an isocost line?

Ans. An isocost line–a tool used in production theory is similar to the budget constraint used in consumer theory. An isocost line is a graph showing all possible combinations of inputs (labour and capital) that can be purchased for a given total cost.

Assuming that a firm uses only two inputs -labour and capital, with price for labour, that is, wage rate, given by w, and price of capital, that is, rental rate or rate of interest, by r, the total cost C is given by

$$C = wL + rK$$

For each different level of total cost, the above equation describes a different isocost line.

To understand the concept more clearly, assume that w is Rs 500 per worker per day and r is Rs 1,000 per unit per day. The following table shows various combinations of capital and labour that would cost the firm Rs 10,000 per day .

Units of capital (at Rs 1,000 per unit)	*Units of labour (at Rs 500 per worker)*	*Total Cost*
10	0	Rs 10,000 = Rs 10,000 + Rs 0
8	4	Rs 10,000 = Rs 8,000 + Rs 2,000
6	8	Rs 10,000 = Rs 6,000 + Rs 4,000
4	12	Rs 10,000 = Rs 4,000 + Rs 6,000
2	16	Rs 10,000 = Rs 2,000 + Rs 8,000
0	20	Rs 10,000 = Rs 0 + Rs 10,000

At one extreme, the producer might spend all of his Rs 10,000 on 10 units of capital Rs 1,000 per unit and have nothing left to spend on labour. Or, by giving up 2 units of capital and thereby saving Rs 2,000, he can hire 8 units of capital at Rs 1,000 each and 4 workers at Rs 500 each. And so on to the other extreme, at which the producer could hire 20 workers at Rs 500 each, spending his entire money on labour input with nothing left to spend on capital.

Rearranging the equation giving isocost line, that is, $C = wL + rK$, we have

$$K = \frac{C}{r} - \frac{w}{r}L$$

And

$$L = \frac{C}{w} - \frac{r}{w}K$$

These two expressions give the quantities of inputs K and L respectively, given the amount of money available for spending C, prices of inputs – w and r, and the quantity purchased of other input.

Suppose the producer decides to use labour input only. It means he uses zero units of capital. Putting the values in the expression

$$L = \frac{C}{w} - \frac{r}{w}K$$

We have

$$L = \frac{\text{Rs } 10{,}000}{\text{Rs } 500} - \frac{\text{Rs } 1{,}000}{\text{Rs } 500} \times 0$$

$$L = 20 - 0$$

$$L = 20$$

It means if the student decides to use labour input only, and given the prices and income, he can hire 20 workers.

Now suppose the producer decides to use capital only. It means he uses zero units of labour input. Putting the values in the expression

$$K = \frac{C}{r} - \frac{w}{r}L$$

We have

$$K = \frac{Rs\ 10,000}{Rs\ 1,00} - \frac{Rs\ 500}{Rs\ 1,000} \times 0$$
$$Z = 10 - 0$$
$$Z = 10$$

It means if the producer decides to use capital only, and given the prices of inputs and money available for spending, the producer can hire 10 units of capital

If we plot the above information in an input space, that is, on a graph whose axis measure quantities of inputs, we will get the following figure

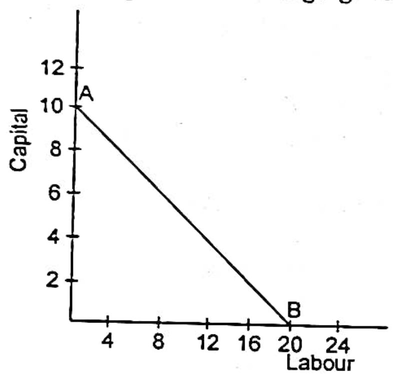

Fig. 8.15

Horizontal axis of the above figure measures labour and vertical measures capital. If the producer decides to spend all his money on labour and nothing on capital, he can hire 20 workers and zero units of capital. This is given by point B.

If the producer decides to spend all his money on capital and nothing on labour, he can hire 10 units of capital and no worker. This is given by point A.

If we join points A and B we get what is called isocost line, the graph which shows the combinations of quantities of the inputs a producer can hire by spending all his given money, given the prices of two inputs.

Points A and B give the extreme cases in which the producer hires only K or only L, respectively. The points between these two points, along the isocost line, represent the other possible combinations of quantities of labour and capital the cost of which must add up to Rs 10,000.

All the combinations of labour and capital on or inside the isocost line are attainable from the Rs 10,000 of money and the given prices. The producer can afford to hire, for example, 8 units of capital at Rs 1,000 each and 4 workers at Rs 500 each. He also can obviously afford to hire 2 units of capitals and 12 workers, thereby using up only Rs 8,000 of the Rs 10,000 available. Thus, attainable combinations of the producer include all the combinations of quantities given by points on the isocost line and below it – the set of combinations of quantities of two inputs that can be hired by spending the whole or a part of the given money. In the diagram, this set is given by the area of triangle AOB.

However, to obtain maximum output, the producer will want to spend the full Rs 10,000. The isocost line shows all combinations that cost exactly the full Rs 10,000.

In contrast, all combinations beyond the isocost line are unattainable. The Rs 10,000 limit simply does not allow producer to hire, for example, 6 units of capital at Rs 1,000 each and 9 workers at Rs 500 each. That Rs 10,500 expenditure would clearly exceed the Rs 10,000 limit. In the figure given above, the attainable combinations are on and within the isocost line; the unattainable combinations are beyond the isocost line.

As with isoquants, a series of isocost lines can be drawn, each one representing a particular cost to the firm. The series of isocost lines is called an isocost map. A hypothetical isocost map is shown in the following figure.

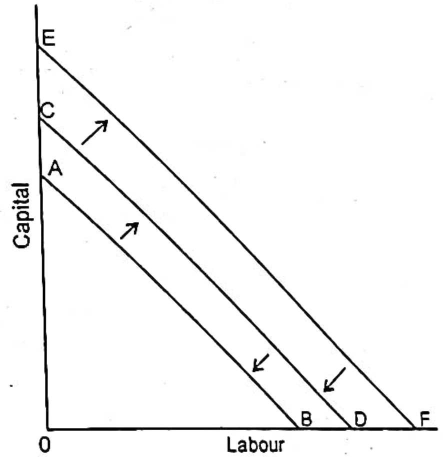

Fig. 8.16

The figure shows graphs of isocost lines for three different total cost levels, TC_0, TC_1, and TC_2, where $TC_2 > TC_1 > TC_0$. In general, there are an infinite number of isocost lines, one corresponding to every possible level of total cost. Moreover, the higher the costs, or the money available for spending, the farther out to the right will the isocost line be.

8. What is the effect of change in prices of two inputs on isocost line?

Ans. Prices of two inputs can change in two ways: either price of only one of the two inputs changes at a time or prices of both inputs change simultaneously. When price of only one input changes at a time, the isocost line shifts but on the axis measuring the input whose price has changed. This is illustrated in the following fig. 8.17.

Suppose AB is the original isocost line determined by certain prices of labour and capital and a certain amount of money available for spending. If the producer spends all his income on labour, he can purchase OL_1 quantity. Suppose the price of labour falls, price of capital and money available for spending remaining unchanged. Now, with a lower price of labour, the producer can hire a higher quantity of labour. Let this higher quantity be given by OL_2, represented by point C under the condition that the producer spends all the money on labour only.

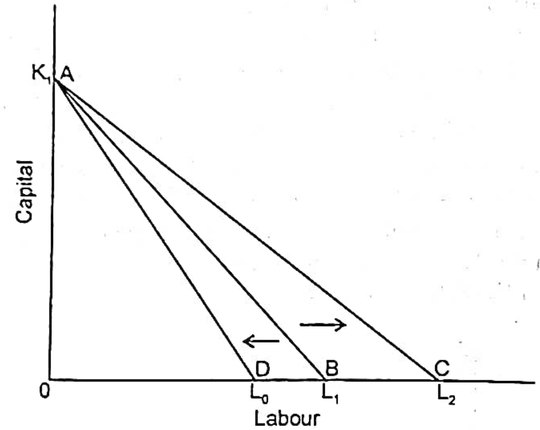

Fig. 8.17

Because price of capital and money do not change, the producer can buy same quantity of capital (given by OK_1) if he spends all his income on capital only even in the new situation of lower price of labour.

If we join point A and point C which gives the new maximum affordable quantity of labour at the new lower price, we get a new isocost line given by AC.

Now suppose the price of labour rises from its original price, price of capital and available money remaining unchanged. With a higher price of labour, the given money can buy a lower quantity of it. Let this lower quantity be given by OL_0 represented by point D under the condition that the producer spends all his money on labour only.

This change in price of labour will again have no effect on the maximum quantity of capital if the producer spends all his money on capital only.

If we join point A and point D which gives the new affordable quantity of labour at the new higher price, we get a new isocost line given by AD.

This illustration shows that when the price of only one input changes, price of other input and amount of money available for spending remaining unchanged, the isocost line shifts on the axis measuring the input whose price has changed: it shifts away from the origin in case of a fall in price of the input indicating that the producer can afford a higher maximum quantity of the input if he spends all his money on this input. The isocost line shifts towards the origin in case of a rise in price of the input indicating that the producer can afford a lower maximum quantity of the input if he spends all his money on this input.

We can repeat the exercise to analyze the effect of change in price of capital on the isocost line, price of labour and money remaining unchanged. The exercise will show that a fall in price of capital will again shift the isocost line away from the origin and a rise in price of capital will once again shift the isocost line towards the origin. But this shift will occur only on axis measuring capital.

If prices of both inputs fall simultaneously so that producer can afford a higher maximum quantity of both the inputs if he spends all his money on either of the input, the isocost line will shift away from origin on both the axis. However, the extent of the shift will

depend on the magnitude of the fall in price: the extent of the shift will be more on the axis measuring the input with a relatively higher fall in price. This is depicted in the following figure on the assumption that prices of both the inputs fall and fall in price of capital is relatively higher.

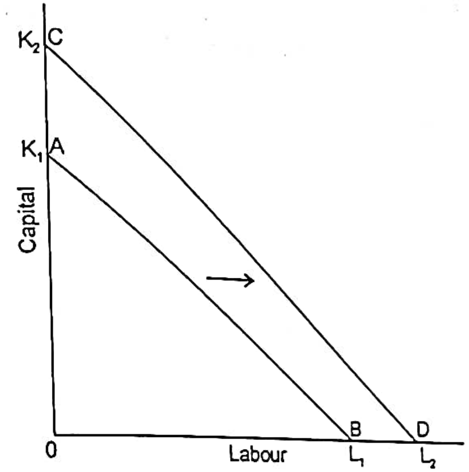

Fig. 8.18

If prices of both inputs rise simultaneously so that producer can buy a lower maximum quantity of both the inputs if he spends all his money on either of the input, the isocost line will shift towards the origin on both the axis. However, the extent of the shift will depend on the magnitude of the rise in price: the extent of the shift will be more on the axis measuring the input with a relatively higher rise in price. This is depicted in the following figure on the assumption that prices of both the inputs rise and rise in price of capital is relatively higher.

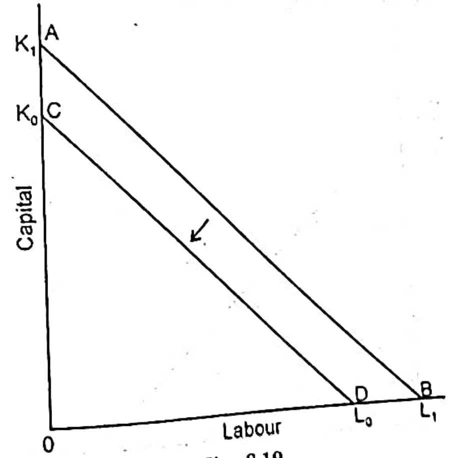

Fig. 8.19

If price of one input rises and that of other input falls, the isocost line will shift away from the origin on the axis measuring the input whose price has fallen and towards the origin on the axis measuring the input whose price has increased so that the new isocost line obtained after the changes in prices have been accounted for will intersect the original isocost line. This is depicted in the following figure which has been drawn on the assumption that price of labour has decreased and that of capital has increased so that budget line shifts from AB to CD.

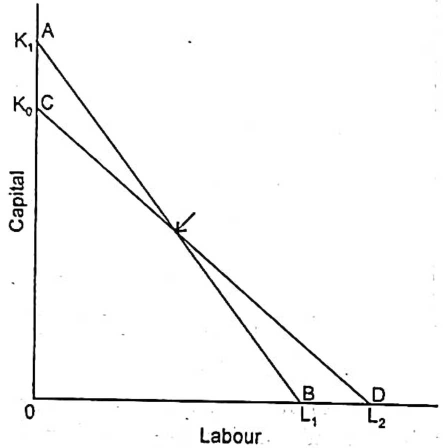

Fig. 8.20

9. Show that the slope of an isocost line represents the ratio of the prices of two inputs.

Ans. The fact that the slope of an isocost line represents the ratio of the prices of two inputs can be shown using the figure given below:

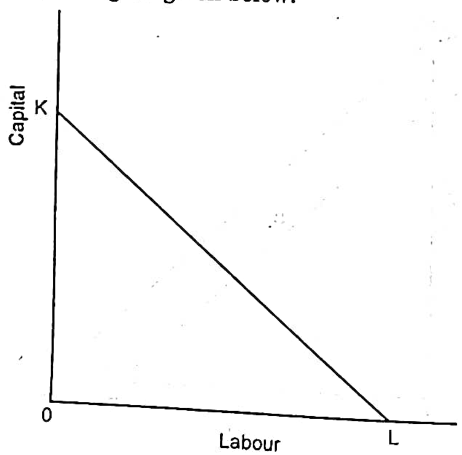

Fig. 8.21

Suppose the amount of money given to producer is C, and w and r are the prices of inputs labour and capital respectively. The slope of the isocost line AB is the tangent of angle ABO which equals $\frac{AO}{OB}$, or simply vertical intercept of the line divided by its horizontal intercept. We intend to prove that the slope of line AB given by $\frac{AO}{OB}$ equals the ratio of the prices of two inputs, w/r. In short, we intend to prove that $\frac{\text{vertical intercept}}{\text{horizontal intercept}} = \frac{w}{r}$.

The vertical intercept is the point at which the producer spends all his money on capital -the input measured on vertical axis, and none on labour -the input measured on horizontal axis, (that is, L = 0). In the above figure this is given by OK. In that case the units of capital the producer hires equal:

$$K = \frac{C}{r} - \frac{w}{r} L$$

$$K = \frac{C}{r} - \frac{w}{r} 0 = \frac{C}{r}$$

In other words, vertical intercept of the line equals $\frac{C}{r}$

Similarly, the horizontal intercept is the point at which the producer spends all his money on labour -the input measured on horizontal axis, and none on capital -the input measured on vertical axis, (that is, K = 0). This is given by OL. In that case the units of labour the producer can hire equal:

$$L = \frac{C}{w} - \frac{r}{w} K$$

$$L = \frac{C}{w} - \frac{r}{w} 0 = \frac{C}{w}$$

In other words, horizontal intercept of the line equals $\frac{C}{w}$.

Now we have the information needed to find the slope of the isocost line. It is:

$$\text{Slope of isocost line} = \frac{\text{vertical intercept}}{\text{horizontal intercept}} = \frac{\frac{C}{r}}{\frac{C}{w}} = \frac{C}{r} \times \frac{w}{C} = \frac{w}{r}$$

It is thus shown that the slope of the isocost line equals the ratio of the prices of the two inputs.

10. Illustrate the choice of optimum factor combination with the help of isocost lines and isoquants.

Ans. We know that a producer aims at maximizing his profits. For this purpose, the producer would like to produce his desired quantity of output at the minimum possible cost. That is to say, a profit maximizing producer will seek to minimise his cost for producing a given output, or to put it in another way, he will seek to maximise his output for a given level of cost.

The input combination which minimises the cost of producing a given level of output is called optimum factor combination or economically efficient factor combination. The process of arriving at the optimum factor combination for a given level of output involves the use of isoquants and isocost lines. Isoquants give the technical possibilities of production, and isocost lines give the cost of production.

In the present scheme of things we have a pre-decided level of output to be produced –y, and given input prices –w and r. The problem is to find that input combination which will involve lowest possible cost for the production of the level of output already decided upon.

The graphic solution to this 'cost minimisation problem' involves the finding of the point on the isoquant giving the already decided upon level of output that has the lowest possible isocost line associated with it. This is shown in the following figure.

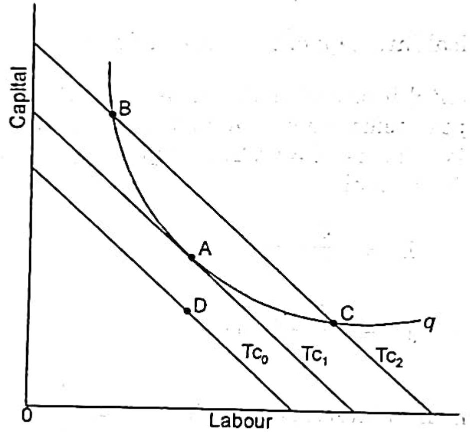

Fig. 8.22

The figure shows three isocost lines each associated with different level of cost depending upon input prices, and an isoquant showing the technically efficient combinations of labour and capital that can produce q level of output.

The solution to the firm's cost-minimization problem occurs at point A, where the isoquant is just tangent to an isocost line. That is, of all the input combinations along the isoquant, point A at which the given isoquant is tangent to an isocost line provides the firm with the lowest level of cost. This is called tangency condition for the solution of cost minimising problem.

To verify this, consider other points in the above figure, such as B, C, D, E and F:

- Point F is off the y isoquant altogether. Although this input combination could produce y units of output, in using it the firm would be wasting inputs (i.e., point G is technically inefficient). This point cannot be optimal because input combination A also produces y units of output but uses fewer units of labour and capital.
- Points B and C are technically efficient, but they are not cost-minimizing because they are on an isocost line that corresponds to a higher level of cost than the isocost line passing through the cost-minimizing point A. By moving from point B to A or from C to A, the firm can produce the same amount of output, but at a lower total cost.
- Point D although on a lower isocost line is associated with a combination of inputs which cannot produce level of output associated with isoquant y.

Note that at the tangency of the given isoquant with an isocost line giving the cost-minimizing point, the slope of the isoquant is equal to the slope of the isocost line, and we know that the negative of the slope of the isoquant is equal to the marginal rate of technical substitution of labour for capital, $MRTS_{L,K}$ and that $MRTS_{L,K} = \frac{MP_1}{MP_K}$. Moreover, we also know that the slope of an isocost line is $\frac{w}{r}$. Thus, the cost-minimizing condition occurs when:

$$\text{Slope of isoquant} = \text{slope of isocost line}$$

$$MRTS_{L,K} = \frac{w}{r}$$

$$\frac{MP_L}{MP_K} = \frac{w}{r}$$

ratio of marginal products = ratio of input prices

The above result can also be written in the following way:

$$\frac{MP_L}{w} = \frac{MP_K}{r}$$

Expressed this way, this condition tells us that at a cost-minimizing input combination, the additional output per dollar spent on labour services equals the additional output per dollar spent on capital services.

11. Illustrate graphically how a firm solves the problem of output maximization for a given level of cost.

Ans. Output maximization for a given level of cost is the dual of cost minimization for a given level of output. In case of cost minimization we have an already decided upon level of output. We need to find out the lowest possible cost with which this given level of output can be produced. In case of output maximization we have an already decided upon level of

cost. We need to find out the highest level of output that can be produced with this given level of cost. In other words, in case of output maximization, given the input prices and the amount of money available to spend, we need to find that input combination which will give us the maximum level of output. This combination of inputs will, once again, be the optimum factor combination, but now for a given level of cost as against optimum factor combination for a given level of output in case of cost minimization.

In the present scheme of things we have a pre-decided upon level of cost –C with input prices – w and r. The problem is to find that input combination which will produce highest possible level of output for the given level of cost.

The graphic solution to this 'output maximisation problem' involves the choosing of the point on the isocost line giving the already decided upon level of cost that has the highest possible isoquant associated with it. This is shown in the following figure.

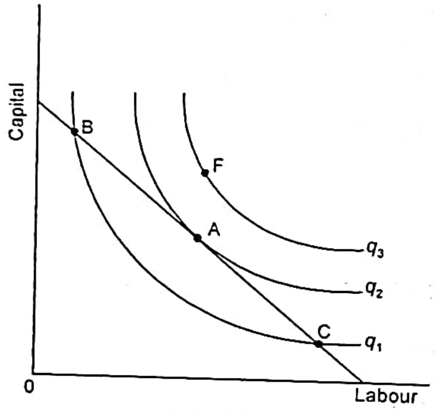

Fig. 8.23

The figure shows three isoquants each associated with different levels of output showing the technically efficient combinations of labour and capital that can produce each level of output, and an isocost line giving the already decided upon level of cost. The solution to the firm's output maximisation problem occurs at point A, where the isocost line is just tangent to an isoquant. That is, of all the input combinations along the isocost line, point A at which the given isocost line is tangent to an isoquant provides the firm with the highest level of output.

To verify this, consider other points in the above figure, such as B, C, D, E and F:

- Point F is off the given isocost line altogether. Although this input combination *could* produce a higher level of output than A, the firm cannot afford this input combination (*i.e.*, point F is economically infeasible or unaffordable).
- Points B and C are economically affordable, but they are not output-maximising because they are on an isoquant that corresponds to a lower level of output than the isoquant passing through the output-maximising point A. By moving from point B to A or from C to A, the firm can produce a higher level of output with the same total cost.

Note that at the tangency of the given isocost line with an isoquant giving the output-maximizing point, the slope of the isoquant is equal to the slope of the isocost line, and we know that the negative of the slope of the isoquant is equal to the marginal rate of technical substitution of labour for capital, $MRTS_{L,K}$, and that $MRTS_{L,K} = \frac{MP_L}{MP_K}$. Moreover, we also know that the slope of an isocost line is $\frac{w}{r}$. Thus, the cost-minimizing condition occurs when:

$$\text{Slope of isoquant} = \text{slope of isocost line}$$

$$MRTS_{L,K} = \frac{w}{r}$$

$$\frac{MP_L}{MP_K} = \frac{w}{r}$$

ratio of marginal products = ratio of input prices

The above result can also be written in the following way:

$$\frac{MP_L}{w} = \frac{MP_K}{r}$$

Expressed this way, this condition tells us that at the output maximising input combination, the additional output per dollar spent on labour services equals the additional output per dollar spent on capital services.

This is the same result we arrived at while solving the cost-minimisation problem of a firm.

12. Illustrate the concept of expansion path.

Ans. In order to minimize its costs for a given level of output or to maximize its output for a given level of cost, a firm needs to arrive at the tangency point between an isoquant and an isocost line. Now the question is, how the firm's production choices and its total costs do change as the optimal production quantity changes. That is to say we want to know how a firm changes its factor combination as it expands its output, given the prices of factors.

We can know this by repeating the procedure for finding the optimal point for many different isocost lines and isoquants. By doing so we will trace out a curve that shows all efficient combinations of two inputs. This is the so called expansion path. An expansion path is a line obtained as a locus of cost minimising points (combinations of inputs) for all possible levels of output, holding input prices constant.

This is illustrated in the following figure.

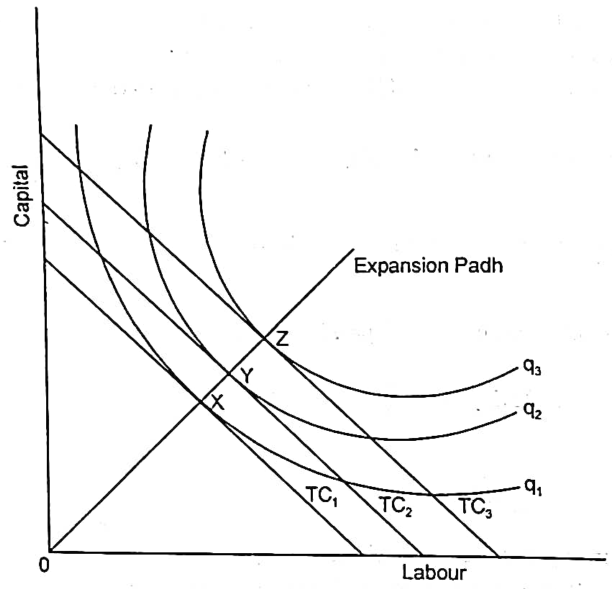

Fig. 8.24

The figure shows sets of isoquants and isocost lines for a hypothetical firm. The figure illustrates only three isoquants and isocost lines but remember that there are isoquants for every possible quantity and isocost lines for every cost level. Note that all isocost lines are parallel to one another indicating that the prices of the two factors remain the same. Shape of isoquants is also same indicating that the technology does not change also.

The combination of inputs (labour and capital) that minimizes the cost of producing a given quantity of output is given at the tangency of an isocost line and the isoquant corresponding to that output level. The figure given above shows three such tangencies. On the lower left, q_1 is the isoquant that corresponds to input combinations that allow the firm to produce q_1 units of output. This isoquant is tangent at point X to the isocost line, TC_1, so TC_1 is the lowest cost at which the firm can produce q_1 units of output. The isoquant representing input combinations that produce q_2 units is tangent to the isocost line TC_2 at point Y, indicating that the firm's minimum cost for producing q_2 units of output is TC_2. At point Z, the q_3 isoquant is tangent to the TC_3 isocost line, so TC_3 is the minimum cost of producing q_3 level of output.

The line connecting the three cost-minimizing input combinations in the above figure (as well as all the other cost-minimizing isoquant-isocost line tangencies for output levels that are not shown) is the firm's expansion path, the line illustrating how the optimal mix of labour and capital varies as the firm expands its level of output.

The expansion path can have different shapes and slopes depending upon the relative prices of the productive factors used, that is, the slopes of isocost lines, and the technical possibilities of producing different levels of output, that is the shape of the isoquants. When the production function exhibits constant returns to scale, the expansion path will be a straight line through the origin. Further, for a given isoquant map, there will be a different expansion path for each different relative prices of the factors.

An important point to note is that the expansion path represents the least-cost combination of inputs to produce a given level of output in the long run, when the firm is able to vary the levels of all of its inputs.

13. Illustrate different types of returns to scale with the help of isoquants.

Before using isoquants to illustrate different types of returns to scale, it will be helpful to recall the meaning of different types of returns to scale. Generally, there are three types of returns to scale: increasing, decreasing and constant.

Increasing returns to scale is a situation in which output rises at a faster rate than the rise in inputs: a doubling of inputs more than doubles the output. Constant returns to scale is a situation in which output rises at the same rate as the rise in inputs: a doubling of inputs doubles the output. Finally, decreasing returns to scale is a situation in which output rises at a slower rate than the rises in inputs: a doubling of inputs less than doubles the output.

In order to illustrate different types of returns to scale with the help of isoquants, we use three diagrams, one for each type of returns to scale. In each diagram, we draw an isoquant map consisting of four isoquants.

In the following figure it is shown that initially the firm uses one worker and one unit of capital, point a. The inputs are doubled consecutively, from a to b to c which lie along the dashed straight line indicating that the proportion in which inputs are combined remains fixed. However, at each doubling of inputs, output more than doubles from $q = 10$ to $q = 30$ to $q = 80$, so the production function has increasing returns to scale.

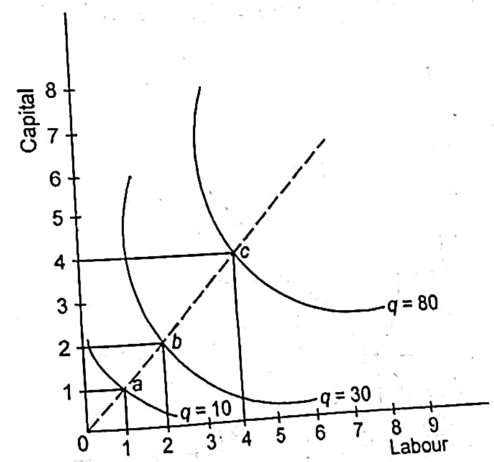

Fig. 8.25

In the following figure it is again shown that initially the firm uses one worker and one unit of capital, point a. The inputs are again doubled consecutively, from a to b to c

which lie along the dashed straight line. However, at each doubling of inputs, output less than doubles from $q = 10$ to $q = 18$ to $q = 34$, so the production function has decreasing returns to scale.

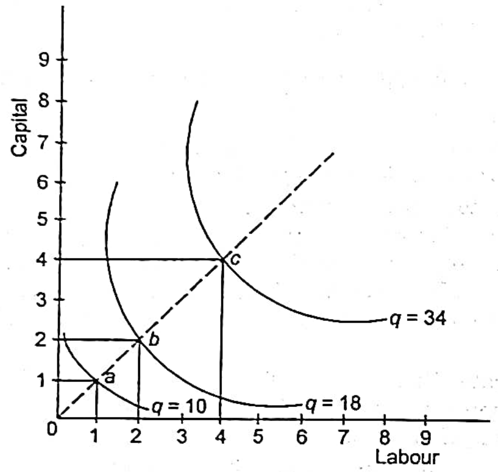

Fig. 8.26

In the following figure it is once again shown that initially the firm uses one worker and one unit of capital, point a. The inputs are, once again, doubled consecutively, from a to b to c which lie along the dashed straight line. In this case, at each doubling of inputs, output also doubles from $q = 10$ to $q = 20$ to $q = 40$, so the production function has constant returns to scale.

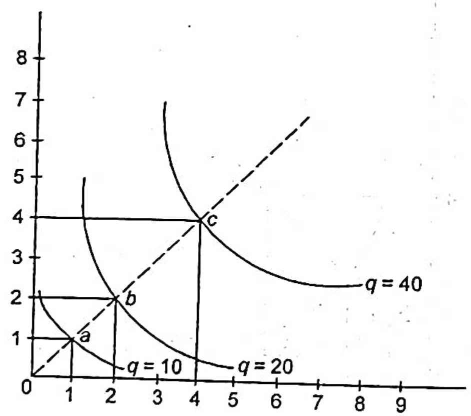

Fig. 8.27

In many industries, returns to scale are estimated to be the same at all levels of output as shown in above diagrams. However, in some cases a production function's returns to

scale may vary as the scale of the firm changes. This means, these production functions have increasing returns to scale for small amounts of output, constant returns for moderate amounts of output, and decreasing returns for large amounts of output. This is shown in the following figure.

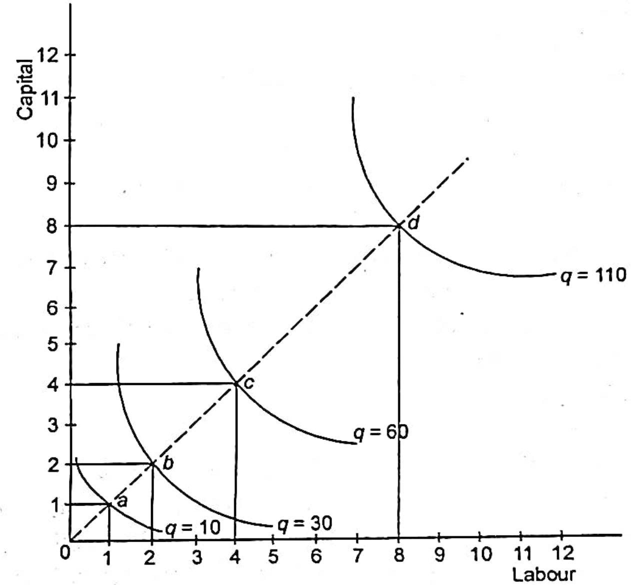

Fig. 8.28

The production function shown above exhibits varying returns to scale. Initially, the firm uses one worker and one unit of capital, point a. It repeatedly doubles these inputs to points b, c, and d, which lie along the dashed straight line. The first time the inputs are doubled, from a to b, output more than doubles from $q = 10$ to $q = 30$, so the production function has increasing returns to scale. The next doubling of inputs, b to c, causes a proportionate increase in output from $q = 30$ to $q = 60$, exhibiting constant returns to scale. At the last doubling of inputs, from c to d, the production function exhibits decreasing returns to scale; inputs double but output less than doubles from $q = 60$ to $q = 110$.

MULTIPLE CHOICE QUESTIONS

1. Isoquants are also called

(A) equal product curves (B) supply curves.

(C) indifference curves. (D) budget constraints.

2. In case of producers, a budget constraint

(A) shows the consumption bundles that a consumer can afford.

(B) reflects the desire by producers to increase their sales.

CHAPTER 10

Economies and Diseconomies of Sales

1. What are economies and diseconomies of scale?

Ans. Economies of scale are the tendencies that may occur as a result of increase in the scale of production that decrease the unit or average cost of production. In other words, economies of scale are the benefits of large scale production resulting in reduced cost per unit of output. The economies of scale are also identified as the situations in which total cost rises at a slower rate than the rise in output.

The cost reductions may accrue either in terms of the reduction in the use of physical quantity of inputs like raw materials, various types of labour and various types of capital per unit of output, that is, increase in the productivity of inputs, or in terms of the payment of lower prices for inputs.

The first set of benefits of large scale production or economies of scale are called real economies of scale. Thus, real economies of scale are the benefits of large scale production associated with the reduction in the use of physical quantity of inputs per unit of output. These benefits accrue to a firm from the factors related to labour, machinery, and management etc. The benefits which are related to labour are called labour economies, those related to machinery or for that matter capital are called technical economies and those related to management are called managerial economies of scale.

The second set of benefits of large scale production or economies of scale are called pecuniary economies of scale. Thus, pecuniary economies of scale are the benefits of large scale production associated with the payment of lower prices on account of bulk-buying of raw materials, increase in bargaining power etc. These economies accrue to a firm from the factors related to selling and marketing, transportation, finance etc. To distinguish them from pecuniary economies, real economies are often referred to as non-pecuniary economies of scale.

Each of the real and pecuniary economies of scale can occur to a firm from the factors which are either internal or external to it. Scale economies which accrue to a firm because of factors which are internal to it are called internal economies of scale and those which accrue to a firm because of factors external to it are called external economies of scale. More precisely, internal economies of scale are those scale economies which accrue to a firm because of the increase in its own scale of production. External economies of scale, on the other hand, are those scale economies which accrue to a firm because of the expansion of the output of the entire industry in which the firm is producing. Internal economies are firm specific: they accrue to that firm only which produces at a larger scale. External economies are available to all firms irrespective of the fact that whether a firm expands its own output or not.

It must be now clear that we have four categories of economies of scale. These are:

1. Internal-real economies of scale
2. Internal-pecuniary economies of scale
3. External-real economies of scale, and
4. External-pecuniary economies of scale

2. Give a detailed account of internal-real economies of scale.

Ans.

Internal- Real Economies of Scale

Internal-real economies of scale are categorised and named on the basis of their sources. These include:

1. Labour Economies : These are the economies of scale which are associated with the use of labour. The main sources of labour economies are:

- Division of labour and specialization, and
- Cumulative volume economies.

Division of Labour and Specialization

Large scale operation allows division of labour and specialization of labour. Division of labour means assigning to each worker a particular task of a productive process. Specialization of labour means concentration by a worker on a particular line of production. Thus, specialization is the result of division of labour.

Division of labour and the resulting specialization lead to improvement in the skills and hence of the productivity of various types of labour. Division of labour also leads to saving of time usually lost in going from one type of work to another. Specialization of labour promotes innovation in the way of doing the work. It also promotes invention of tools and machines helpful to workers. All these factors result in the increase in productivity of labour and hence to increasing returns to scale.

Cumulative Volume Economies : With increasing scale there is also an experience effect on the skills of employees, where experience is measured as the sum of all past outputs. As the volume of production increases, employees tend to do greater number of operations. The result is that they acquire cumulative or increasing experience leading to higher productivity and hence to more than proportionate increase in output. This effect is also called learning by doing or economies of learning.

2. Technical Economies : These are the economies of scale associated with the use of fixed capital, that is, machinery and equipment. These are also called economies of mass production because as the scale of a firm's operation expands, it can begin to utilize large-scale machines and production systems that can substantially reduce cost per unit. The main sources of technical economies are

- Indivisibilities of capital equipment
- Specialization of capital
- Technical volume-input relations, and
- Reserve capacity

Indivisibilities of Capital Equipment

Indivisibility or lumpiness of factors inputs means that some factors inputs are available in some minimum sizes; they are not available below that minimum size and cannot be divided into smaller sizes to suit a smaller level of production. Examples of such factors include machinery, buildings (also called factory space), and land. In other words, these factors can be acquired or used only in certain finite sizes and in some cases are more efficient in larger than in smaller sizes.

In the beginning when the level of production is small these factors are not used fully, there is incomplete or under-utilisation of such lumpy factors which results in higher per unit cost associated with these factors. As the scale of production increases, there is fuller and intensive, hence more efficient use of these factors which results in higher productivity and in the reduction in unit cost of production associated with such factors.

Take the case of a big tractor. A small-scale farmer cannot make full use of one. They only become economical to use on farms above a certain size. As the farm size increases they are used more efficiently and the tractor related cost goes on decreasing.

Specialization of Capital

At a small output rate, firms often must employ general purpose machines and tools. As the size of operations increases, it can use more specialised and sophisticated machinery there by lowering its cost. For example, with an output of a few litres an hour, a fruit-juice making company must use regular blender like the ones in our kitchens. But if the company increases its production to hundreds of litres an hour, it has to use commercial blenders that fill, empty, and clean themselves. The result is that the output rate is larger and the average total cost of producing a litre of fruit-juice is lower.

Technical Volume-Input Relations

These are also called dimension effects and technical geometric relationships between particular equipment and the inputs required to produce it. The idea is that when the dimensions of certain types of plant such as ware houses, storage tanks etc are increased, their capacity increases more proportionately than the increase in their cost of construction. It is a geometric fact that the material and labour used in constructing such type of plant as connecting pipes, ware houses etc are proportional to the surface area that they occupy. But the volume capacity (which determines the volume of output) of a plant increases more than proportionately as the area increases. Hence, the technical or physical cost of constructing such industrial plants falls as the output capacity increases.

Reserve capacity

Firms always want some reserve capacity in order to avoid disruption of the production process because of the break-down of machinery. A small firm which uses a single large machine will have to keep two such machines if it wants to avoid disruption in the production process from a breakdown. A larger firm which uses several large machines can attain the required security by holding a smaller proportion of this total number as reserve capacity.

3. Managerial Economies : Economies of scale associated with functions of managerial staff are called managerial economies. These economies arise on account of following three factors:

- Specialisation in the managerial functions
- Decentralisation of decision making, and
- Mechanisation of managerial functions

Specialisation in the Managerial Functions

When the size of a firm becomes large, the division of managerial functions becomes possible. For example, in a small firm a single person takes decisions regarding, production, sales, finance, advertising, hiring of employees and other such matters. But in a large firm these matters are handled separately by separate persons, for example, production manager, sales manager, finance manager, personnel manager, marketing manager etc. This results in specialisation in managerial functions which in turn increases the productivity of the managerial staff.

Decentralisation of Decision Making Process

Assigning different managerial tasks to different persons, which is possible only in a larger firm leads to decentralisation of decision making process. Decentralisation of decision making process reduces many types of delays and disturbances in the flow of managerial information resulting in an increase in overall efficiency and hence productivity.

Mechanisation of Managerial Functions

Large firms also apply techniques of management involving a high degree of mechanisation such as telephones, telex machines, television screens, computers etc. These techniques save the time in the decision making process and speed up the processing of information as well increasing its amount and accuracy. This becomes another source of higher productivity and hence reduction in unit cost of output.

3. Give a detailed account of internal-pecuniary economies of scale.

Ans. Pecuniary economies of scale are the benefits of large scale production associated with the payment of lower prices on account of bulk-buying of raw materials, increase in bargaining power etc. some of the pecuniary economies of scale internal to a firm are:

1. **Advantages of Bulk-Buying :** Bulk-buying means purchasing something in large quantities. When a firm attains a sufficiently large size, it needs large quantities of raw materials. The suppliers of these raw materials often provide attractive discounts to these bulk-buyers. This reduces the prices of raw materials needed by such large size firms.
2. **Advantages of Increased Bargaining Power :** Even if suppliers are reluctant to give discounts, large size gives a firm an additional advantage of increased bargaining power. Because a large firm needs larger quantities of raw materials, it can bargain with a supplier with more power than a firm of smaller size. The result is that it can obtain raw materials at reduced prices.

Large size of a firm does not give it increased bargaining power for negotiating prices of raw materials only, it can get an advantage in negotiating wages and salaries of workers also, and there are high chances that a large size firm can succeed in negotiating lower wages and salaries with workers than a firm with smaller size.

3. **Benefits of Prestige :** Large size and well known firms have prestige associated with their employment. It means workers consider it to be a matter of prestige to work with a large and well known firm. In order to get employment in such a firm, workers may not demand as high salaries as they would in a smaller and less known firm. This also gives large size firms the benefit of paying lesser wages and salaries.
4. **Savings on Advertisement Cost :** Large firm often need to advertise at large scales. This may tempt advertisers to grant lower prices to such large firms in order to attract these large size advertising orders or deals.
5. **Savings in Transportation Costs :** Large size firms are required to transport large amounts of raw materials and finished goods. Because special frights are often granted for larger quantities transported, large size gives firms the opportunity to obtain lower rates for transportation.
6. **Advantages in obtaining External Finance at Reduced Prices :** Because large size firms are considered less risky and more credit worthy, banks usually offer loans to them at lower rates of interest and other favourable terms. For example, for having a bank overdraft facility, a supermarket may be charged 2 or 3 % less than a small independent retailer.

4. What are internal diseconomies of scale?

Ans. Diseconomies of scale are the tendencies that may occur as the result of increases in the scale of production to very large sizes that decrease the productivity of factor inputs and hence increase the unit cost of production. In other words, diseconomies of scale are the costs or disadvantages resulting in higher cost per unit of output which accrue to a firm because of production of the good it produces being undertaken on a very large scale or size. Diseconomies of scale are also identified as the situations in which total cost rises faster than the rise in output.

Internal diseconomies of scale are diseconomies of scale which accrue to a firm because of it own actions.

Some of the sources of internal diseconomies of scale are:

1. **Congestion and Overcrowding :** When the scale of operations continues to increase and becomes very large, factor inputs, particularly labour input, gets over crowded or congested. This hinders and obstacles the smooth functioning of factor inputs affecting their productivity adversely.
2. **Managerial Diseconomies :** The most important cause for diminishing returns to scale is found in diminishing returns to management or administrative organisations. It is argued that as the scale of operations becomes very large, management becomes very complex, managers are overworked and the management process becomes less efficient. Even when authority is delegated to individual managers, the final decisions have to be taken by the final centre of top management (Board of Directors). As the output grows to very large sizes, top management eventually becomes over burdened and less efficient in its functioning as coordinator and ultimate decision maker.

Moreover, as the scale of operations becomes very large, top management losses its control. As the layers of managers increase, decisions are delayed in the bureaucracy of large size firms. Also, as the layers of managers increase, information is consciously or unconsciously distorted as it passes through the various hierarchical levels of management. There are also the possibilities of inflexible regulations and red tape. All these factors tend to affect the efficiency of management adversely which becomes an important source of diseconomies of scale.

3. **Communication Problems :** Larger firms often suffer poor communication because they find it difficult to maintain an effective flow of information between departments, divisions or between head office and subsidiaries. Time lags in the flow of information can also create problems in terms of the speed of response to changing market conditions. For example, a large supermarket chain may be less responsive to changing tastes and fashions than a much smaller, 'local' retailer.
4. **Coordination Problems :** Co-ordination problems also affect large firms with many departments and divisions, and may find it much harder to co-ordinate its operations than a smaller firm. For example, a small manufacturer can more easily co-ordinate the activities of its small number of staff than a large manufacturer employing tens of thousands.
5. **Drying up of Advantages of Bulk-buying :** The advantages of paying a lower price from buying in bulk may disappear once certain quantities are reached. At some point, available supplies of key inputs may be limited, pushing their costs up.
6. **Performance Deterioration :** The chances of performance deterioration, also called system slack, are more in larger firms than in smaller ones. There can be many reasons for the deterioration in the performance of workers. Some of these are; a feeling of alienation among workers from their employees; greater opportunities to shirk, in favour of on-the-job leisure; surfacing of laziness among workers because of the ease to hide it, etc.

7. What are external economies and diseconomies?

Ans. External Economies : External economies also called 'spill overs', are those scale economies which accrue to a firm because of the expansion of the output of the entire industry in which the firm is producing As an industry grows in size, each of its member firm, whatever its own individual size, benefits from the whole industry getting larger. External economies are available to all member firms of an industry irrespective of the fact that whether a firm expands its own output or not.

Main sources of external economies include

1. **Benefits from Technological Improvements :** When an industry expands, it may lead to discovery of new technical knowledge. This may happen for the reason

that specialist firms engaged in research and development activities related to the industry may come up. This will definitely make possible the use of improved and better quality machines than before.

2. **Benefits of Skilled Labour :** As an industry grows, workers are attracted towards it. In order to work in the growing industry, workers need to learn the skills needed in the growing industry. Some firms specialising in training of job seekers in the industry may come up. As a result, a pool of trained labour equipped with the traditional skills needed in the industry is made available. This has a favourable effect on the productivity of an individual firm.

3. **Growth of Correlated Industries :** As an industry grows, it will generate waste in large amounts. Some specialised firms may come into existence to process the waste generated by the industry into useful products. Because of this development, the firms of the industry can sell their waste at good prices. This will reduce their cost of production.

4. **Availability of Cheaper Materials :** As an industry grows, increasing the demand for raw materials, tools, equipment etc used by the industry, some firms may come up specialising in the production of raw material, tools, equipment, machinery etc. These firms will produce them on a large scale. This production at large scale by specialist firms will reduce the cost of production of these raw materials, tools etc enabling firms to obtain them at lower prices.

5. **Benefits of Improved Infrastructural Facilities :** When an industry grows in size, it attracts investments in areas like finance, communication, transport, storage, energy, support services etc. This is the industry's infrastructure. As these infrastructural facilities get developed, the member firms of the industry start benefiting as they enjoy cost reductions.

6. **Benefits from Improved information :** As an industry grows in size, many scientific and trade journals and magazines relating to the industry's business are published. These journals provide information relating to markets, sources of raw materials, latest techniques of production etc. The information of this sort is very helpful to the member firms in improving their production techniques and marketing strategies.

7. **Benefits from Advertisements :** When an industry grows in size particularly because of an increase in the number of its member firms, the volume of advertisements of the products of this industry grows very much. This is very beneficial to the member industries as it acts as an attraction for the customers from other related industries.

Among the above mentioned external economies of scale benefits from technological improvements, benefits of skilled labour, growth of correlated industries, benefits from improved infrastructural facilities, and benefits from improved information are real external economies of scale, and availability of cheaper raw materials and benefits from advertising are pecuniary external economies of scale.

External Diseconomies : The member firms of a particular industry might experience external diseconomies of scale also. External diseconomies are the unfavourable effects of

the increase in the size of an industry on its member firms which result in the rise in unit cost of production of these firms.

One source of external diseconomies is congestion. Shortage of specific raw materials or skilled labour is another source of external diseconomies.

1. **Exhaustion of Natural Resources :** Another cause of decrease in productivity of factor inputs as scale of production is increased to very large size or decreasing returns to scale in some industries may be found in exhaustible natural resources. For example, doubling of mining apparatus may not double the extraction of mineral ores.

2. **Shortage of Specific Inputs :** Related to the exhaustion of natural resources is the shortage of some specific inputs like critical raw material and skilled labour. The concentration of similar firms in an area may lead to an increase in demand for inputs like raw materials and labour used by the firms. The competition among firms for these inputs will cause the prices of inputs to increase provided their supply remains unchanged. This consequently would increases the cost of production in the industry.

3. **Urbanization Problems :** The localization of firms in an area results in urbanization problems such as traffic congestion. This slows down the movement of labour, goods and raw materials and thereby retarding the rate at which goods and services are produced. This increases the cost of production in the industry.

4. **Problems of Waste Disposal :** As an industry grows to a large size in terms of number and size of firms, problems of waste disposal may arise. As firms undertake production in this situation, large amounts of waste are bound to be generated. Firms may be compelled to employ costly waste disposal methods in order to keep the area clean.

5. **Increased Expenditure on Sales Promotion :** As the number of firms in an industry increases, individual firms will have to face increased competition from the increased number of their rivals. As a result increased expenditure on sales promotion through advertisement etc would have to be resorted to and more money will have to be spent on that if each firm is to maintain its position.

Among the above mentioned external diseconomies of scale, exhaustion of natural resources, shortage of specific inputs and increased expenditure on sales promotion may be regarded as pecuniary external diseconomies of scale, and urbanisation problems and problems of waste disposal may be regarded as real external diseconomies of scale.

8. What is the relation between economies of scale and returns to scale?

Ans. Economies of scale and returns to scale are not the same thing. They are related—cost and the level of inputs move closely together—but there is a difference. Returns to scale describe how output changes when all inputs are increased by a common factor. But there is nothing which prevents cost-minimizing firms from changing the input ratios when they increase output. When input proportions change, the firms expansion path is no longer a straight line, and the concept of returns to scale no longer applies.

The measure of economies of scale, which is about how total costs change with output, does not impose the restriction of constant input ratios the way returns to scale does. Because a firm can only reduce its cost more if it is able to change its input ratios when output changes, it can have economies of scale if it has constant or even decreasing returns to scale. That is, even though the firm might have a production function in which doubling inputs would exactly double output, it might be able to double output without doubling its total cost by changing the proportion in which it uses inputs. Therefore, increasing returns to scale imply economies of scale, but not necessarily the reverse. It is helpful to compare the two:

Increasing Returns to Scale :	***Output more than doubles when the quantities of all inputs are doubled.***
Economies of scale :	***A doubling of output requires less than a doubling of cost.***

The only case in which the two concepts are the same is when it happens to be optimal for the firm to hold input ratios constant as output increases. (This would show up as an expansion path that is a straight line extending out from the origin.) In this case, the firm gets no extra cost reduction from changing the proportions in which it uses its inputs as its output changes. It means returns to scale is a special case of economies of scale.

As an *industry* grows in size, this can lead to *external economies of scale* for its member firms. This is where a firm, whatever its own individual size, benefits from the *whole industry* being large. For example, the firm may benefit from having access to specialist raw material or component suppliers, labour with specific skills, firms that specialise in marketing the finished product, and banks and other financial institutions with experience of the industry's requirements. What we are referring to here is the *industry's infrastructure*: the facilities, support services, skills and experience that can be shared by its members.

The member firms of a particular industry might experience *external diseconomies* of scale. For example, as an industry grows larger, this may create a growing shortage of specific raw materials or skilled labour. This will push up their prices, and hence the firms' costs.

One source of external diseconomies is congestion. The airline market provides a good example. With bigger airline market output, congestion at both airports and in the air increases, resulting in longer delays and extra waiting time for passengers and airplanes. These external diseconomies mean that as the output of air transportation services increases (in the absence of technological advances), average cost increases. As a result, the long-run market supply curve is upward sloping. A permanent increase in demand brings an increase in quantity and a rise in the price. (Markets with external diseconomies might nonetheless have a falling price because technological advances shift the long-run supply curve downward.)

An example of external economies (spill overs) is the growth of specialist support services for a market as it expands. As farm output increased in the nineteenth and early twentieth centuries, the services available to farmers expanded. New firms specialized in the development and marketing of farm machinery and fertilizers. As a result, average farm

costs decreased. Farms enjoyed the benefits of external economies. As a consequence, as the demand for farm produce

Why does southern California have a comparative advantage in making movies or Switzerland in making watches or New York in providing financial services? The answer is that once an industry becomes established in an area, firms that locate in that area gain advantages over firms located elsewhere. The advantages include the availability of skilled workers, the opportunity to interact with other firms in the same industry, and proximity to suppliers. These advantages result in lower costs to firms located in the area. Because these lower costs result from increases in the size of the industry in an area, economists refer to them as external economies. External economies Reductions in a firm's costs that result from an increase in the size of an industry.